Natural single amino-acid polymorphism (F19Y) in human galectin-8: detection of structural alterations and increased growth-regulatory activity on tumor cells

Federico M. Ruiz[1#], Barbara A. Scholz[2#], Eliza Buzamet[3#], Jürgen Kopitz[4], Sabine André[2], Margarita Menéndez[3], Antonio Romero[1*], Dolores Solís[3*] and Hans-Joachim Gabius[2*]

[1] Departamento de Biología Físico-Química, Centro de Investigaciones Biológicas, CSIC, E-28040 Madrid, Spain

[2] Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität München, D-80539 München, Germany

[3] Departamento de Química-Física Biológica, Instituto de Química Física Rocasolano, CSIC, and Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), E-28006 Madrid, Spain

[4] Abteilung Angewandte Tumorbiologie, Universitätsklinikum Heidelberg, D-69120 Heidelberg, Germany.

Running title: F19Y Galectin-8 variant

Keywords

agglutination, β-sandwich, carcinoma, ganglioside, lectin, proliferation

Correspondence

[*] To whom correspondence should be addressed (E-mail: romero@cib.csic.es or d.solis@iqfr.csic.es or gabius@tiph.vetmed.uni-muenchen.de)

[#] These authors contributed equally to this work

Abbreviations

CRD, carbohydrate recognition domain; CD, circular dichroism; hGal-1, human galectin-1; hGal-8, human galectin-8; ITC, isothermal titration calorimetry; MMP, matrix metalloproteinase; SNP, single-nucleotide polymorphism

Abstract

Natural amino-acid substitution by single-site nucleotide polymorphism can become a valuable tool for structure-activity correlations, especially if evidence for association to disease parameters exists. Focusing on the F19Y change in human galectin-8, connected clinically to rheumatoid arthritis, we here initiate the study of consequences of a single-site substitution in the carbohydrate recognition domain of this family of cellular effectors. We apply a strategically combined set of structural and cell biological techniques for comparing properties of the wild type and variant proteins. The overall hydrodynamic behavior of the full-length protein and of the separate N-domain is not noticeably altered, but displacements in the F0 β-strand of the β-sandwich fold in the N-domain are induced, as evidenced by protein crystallography. Analysis of thermal stability by circular dichroism spectroscopy revealed perceptible differences for the full-length proteins, pointing to an impact of the substitution beyond the N-domain. In addition, small differences in thermodynamic parameters of carbohydrate binding are detected. On the level of two types of tumor cells, characteristics of binding appeared rather similar. In further comparison of influence on proliferation, the variant proved to be more active as growth regulator in the six tested lines of neuroblastoma, erythroleukemia and colon adenocarcinoma. The seemingly subtle structural change identified here thus has functional implications *in vitro*, encouraging further analysis in autoimmune regulation and, in a broad context, work with other natural single-site variants, using the documented combined strategy.

Introduction

The growing awareness of the unsurpassed coding capacity of carbohydrates provides increasing incentive to study the expression and structures of mammalian lectins [1]. These sugar-binding proteins translate glycan-encoded messages into responses on the level of cells [2]. By targeting distinct β-galactosides of selected glycoconjugates, the members of the galectin family do so, the formation of a lectin-glycan complex then initiating signaling, for example for growth regulation, or facilitating glycoprotein delivery to distinct cell types or within apical or axonal transport [2-4]. Obviously, these effectors are an attractive object to study the molecular route from coding by glycans to function.

A common way to delineate structure-function correlations for proteins is the strategic introduction of single-site mutations. For human galectins, the essential nature of a sequence signature, especially a strictly conserved Trp residue for ligand contact, could hereby be documented, starting with work on human galectin-1 (hGal-1) [5]. Later, long-range effects that alter thermodynamic parameters of ligand binding were detected, explicitly for the C2S and R111H mutants of this lectin [6]. These studies underline the significance of work with engineered variants, revealing that even small sequence changes can alter protein properties beyond the site of substitution. Considering the natural occurrence of single-

nucleotide polymorphisms (SNPs), the comparative analysis of the resulting mutant proteins offers an attractive approach for structure-activity research. Of particular note, these deviations from the common genotype may show correlations to clinical parameters, giving work on the variant proteins a biomedical dimension. Herein, we initiate the structural and functional characterization of SNP-based variation in the carbohydrate recognition domain (CRD) of galectins.

Up to now, SNP identification in this lectin family mostly concerned other regions of the genes. For galectins-1 and -2, these are the promoter as well as 5'- and 3'-flanking sequences and introns 1 and 2, with indications for an association to autoimmune myasthenia gravis (hGal-1) and myocardial infarction (hGal-2) [7-10]. As to SNP presence in the coding region, two allelic variations had been discovered in the sequence section for collagen-like repeats of chimera-type galectin-3, at positions 191 (P64H) and 292 (T98P) [11]. The P64H substitution, only found in humans, confers resistance to cleavage by matrix metalloproteinases (MMPs), especially MMPs-2/-9, on the level of the protein, and to doxorubicin for cells, increases sensitivity to death receptor-mediated apoptosis and appears to be associated with increased incidence of breast cancer but reduced risk for prostate carcinoma [12-14]. Occurrence of the T98P change has been reported to be increased in Taiwanese patients with rheumatoid arthritis and Chinese glioma patients as well as related to shortened survival in Chinese non-small cell lung cancer patients [15-17]. In contrast the P64H substitution showed no correlation to disease onset or progression in these three studies. Recently, four non-synonymous SNPs in exons of the gene for human galectin-8 (hGal-8) have been pinpointed. SNP rs2737713 (A>T, which causes a F19Y substitution) was selected for quantitative trait-association study and found to be significantly associated with rheumatoid arthritis [18]. Since this is the first SNP hitting a galectin's CRD examined, which is related to a clinical parameter, we addressed the question on alterations of structure and cell biological functions using an interdisciplinary approach.

Gal-8 is a tandem-repeat-type protein, in which two different CRDs are connected by a linker [19]. Phylogenetically, its representation in the galectin network is rather conserved; e.g. the chicken genome harbors an orthologue as the only tandem-repeat-type galectin [4, 20]. It is widely present in human tissues, localized extra- and/or intracellularly (cytoplasm and/or nucleus), secreted from cells, and useful for diagnostic and prognostic assessments in histopathology, with an expression profile different from those of other galectins [21-26]. As cellular effector, it induces apoptosis or growth arrest in immune (synovial fluid cells of rheumatoid arthritis patients, $CD4^{high}CD8^{high}$ thymocytes) and tumor cells, and also modulates cell adhesion, here by virtue of reactivity with integrins such as the $\alpha_3\beta_1$-heterodimer and integrin counterreceptors [27-31]. A tangible role in autoimmune regulation is further substantiated by its concentration-dependent activity to promote T cell proliferation without antigen (high concentration) and to act as co-stimulator together with antigen (low concentration) [32, 33]. Moreover, it assists Gal-1 in

differentiating mature B cells into plasma cells [34]. Fittingly, sera of patients with autoimmune disease contain autoantibodies against this galectin, whose production in patients with systemic lupus erythematosus reached the status of an independent biomarker for secondary anti-phospholipid syndrome [35, 36].

On the level of glycans as ligands, 3'-sialylated/sulfated β-galactosides, N-acetyllactosamine repeats and histo-blood group ABH epitopes are preferred binding partners of the lectin, with affinity differences between the N- and C-domains, when tested separately [37-42]. Crystallographic analyses of the N-domain pinpointed Arg-59 as the most critical amino acid for the up to nM affinity toward 3'-sialylated/sulfated lactosides, and its absence in the C-domain explained the marked disparity in the respective affinity when examining both domains [43, 44] and also galectin-9's N-domain [45]. Recently, the crystal structures of a linker-trimmed variant of hGal-8 and of the separate C-domain in complex with a peptide of the autophagy cargo receptor NDP52 have become available [46, 47]. The carbohydrate specificities of the two CRDs and their mode of presentation as pair connected by the linker determine Gal-8's functionality as lectin. Occurrence of the SNP (in exon 1), mentioned above, may affect one or both aspects, a question answerable by structural and functional studies. Toward this end, we have devised a strategic combination of different approaches. We first report on the hydrodynamic behavior, circular dichroism (CD)-based analysis of structure and thermal stability of wild type and variant proteins, followed by presenting the crystal structures of the N-domains, then a detailed thermodynamic analysis of ligand binding, and finally function-oriented results on galectin binding to cells, regulation of growth in six different tumor cell lines and mediation of aggregation of erythrocytes.

Results

Quaternary structure and hydrodynamic properties

We started the comparative analysis of wild type hGal-8 and its F19Y mutant, working in parallel with the sets of full-length proteins and the N-domains, by characterizing their behavior in sedimentation equilibrium in the absence and presence of 0.1 M lactose. Data obtained at loading concentrations in the range of 0.3-1.3 mg/ml could invariably be fitted to a model with a single ideal component with weight-average molecular masses comparable to the theoretical masses of the monomers and to those experimentally determined by mass spectrometry (Table 1). The experimental values were independent of rotor speed, indicating homogeneity of protein preparations. Thus, in this range of concentrations, all the proteins behaved in solution predominantly as monomers.

Gel filtration analysis was performed in the presence of 0.1 M lactose to block any carbohydrate-dependent interaction with the resin. All the proteins eluted as a single sharp peak with no evidence for aggregate formation. Apparent molecular masses calculated from the elution times (Table 1) were in very

good agreement with monomer theoretical masses, showing that the hydrodynamic features of the full-length proteins and separate N-domains do not deviate significantly from those expected for globular proteins. Sedimentation velocity experiments further supported gel-filtration and sedimentation equilibrium results. The proteins were found to migrate as a main peak with almost identical $s°_{20,w}$ values in the absence and presence of lactose (Table 1). This result excludes major changes in the hydrodynamic shape of the proteins by ligand loading, which had been encountered for hGal-1 as reduction of the gyration radius or increase of the diffusion constant [48, 49]. Frictional coefficient ratios (f/f_o) of about 1.1 for the separate N-domains and 1.35 for the full-length proteins were calculated. In comparison, the frictional ratio for human hGal-1 was 1.3 [50], suggesting that there are no main differences in the overall hydrodynamic shape between these two tandem-repeat-type proteins and the proto-type galectin, apart from a predictable mass-proportional alteration of the length.

Overall, the results of the gel filtration and ultracentrifugation analyses documented that the F19Y mutation does not have any traceable impact on the association state or hydrodynamic properties of either the separate N-domains or the full-length proteins. Besides, they indicated that dimers of full-length hGal-8 or its N-domain detected by chemical cross-linking at 5 μM protein concentration in the presence of a 50-fold molar excess of bis(sulfosuccinimidyl) suberate [40] can be interpreted as covalent trapping of rather transient contacts. In the quest to detect any differences between wild type and mutant proteins CD spectroscopy was next applied.

CD-based analysis of the structure and thermal stability in solution

The far-UV CD spectrum of hGal-8 showed the presence of a negative band centered at 218 nm (Fig. 1A), consistent with β-sheet structure. No differences were detected in the spectra of the wild type and F19Y proteins. When normalized per mole of molecule, the ellipticity signals for the full-length proteins approximately doubled those registered for the separate N-domains, in agreement with their expected total content in β-sheet structure. Interestingly, a slightly higher negative ellipticity in the 215-225 nm region was detected for the wild type N-domain compared to its F19Y mutant, hinting at a potentially distinctive feature of the β-sandwich arrangement of this domain resulting from the Phe to Tyr substitution. When monitored in the presence of 0.1 M lactose, spectra obtained for all the proteins maintained their profiles, indicating that there were no significant changes in secondary structure upon ligand binding (data not shown).

The near-UV CD spectrum of hGal-8 was characterized by the presence of a broad positive signal, from 260 to 300 nm, with well-defined bands at 264, 272, 282 and 291 nm (Fig. 1B), attributable to fine structure of Phe, Tyr and Trp residues. The spectrum was sensitive to the F19Y substitution, with a noticeable decrease in intensity of the peaks at 272 and 282 nm. Differences were proportionally higher when comparing the spectra of the N-domains, most probably due to the greater relative contribution of

residue 19 (Phe *vs* Tyr) to the overall spectrum of the molecule. In all cases, a significant increase in the ellipticity signals at 282 and 291 nm, in the tyrosine/tryptophan region, was observed when the spectra were obtained in the presence of lactose, while changes in the 264-272 nm region were very minor or not even present (Fig. 2A-D), suggesting that the environment of phenylalanine residues is only barely perturbed by lactose binding. However, for all the proteins, loading with 3'-sialyllactose affected the whole spectrum, including the region attributable to phenylalanine residues (Fig. 2A-D), revealing that binding of this ligand has a larger impact than lactose on the environment of aromatic residues and/or protein dynamics. Whether the noticed differences in the near-UV CD spectra are translated into different properties of the wild type and mutant proteins was first examined by measuring their thermal stability in the absence and presence of ligands.

Stepwise heating of all the proteins (full-length and N-domains) resulted in the loss of secondary structure, as evidenced by a progressive decrease in the intensity of the far-UV negative band. Therefore, the denaturation process was monitored by measuring the change in ellipticity at 218 nm as a function of temperature. A phenomenological analysis of the curves using a sigmoidal function yielded the $T_{1/2}$ values compiled in Table 1. Of note, the full-length proteins apparently denatured in a single cooperative process (Fig. 1A, *inset*), as also observed for the separate N-domains. However, while the thermal stability of the N-domains was very similar, the $T_{1/2}$ of the full-length form of the F19Y variant was 2.5 ºC higher than that of the wild type protein. This result identifies an impact of the F19Y substitution on the level of the full-length protein not confined to the N-terminal domain.

In all cases, thermal stability increased with lactose binding (Fig. 1A, *inset* and Table 2). $T_{1/2}$ values were 9.1 to 9.5 ºC higher for the N-domains and wild type hGal-8, whereas the increase for the F19Y variant was 7.4 ºC, again unveiling distinctive features for this protein. In view of the differences in near-UV CD spectra run in the presence of 3'-sialyllactose, a similar analysis of the effect of protein loading with 3'-sialyllactose was performed. It yielded a different picture. In particular, the stabilization induced by this ligand was similar for wild type and F19Y hGal-8 but the $T_{1/2}$ increase was considerably higher (more than 6 ºC) for the separate N-domains than for the full-length proteins. As only the N-terminal CRD is responsible for 3'-sialyllactose binding and the affinities are similar for all proteins (please see below), the lower degree of 3'-sialyllactose-induced stabilization of the full-length forms should be related, again, with denaturation features of regions beyond the N-domain. Having herewith delineated significant differences in thermal stability, a meticulous examination of possible alterations at the atomic level was performed using X-ray crystallography. Attempts to crystallize the full-length proteins were not successful. In fact, no crystal structure of a full-length mammalian tandem-repeat-type galectin has been solved to date, only the structure of a non-physiological linker-trimmed hGal-8 variant having been reported [44]. Thus, we focused on the N-domains for comparative analysis, as described below, the

position of the sequence variation being suited for this processing.

Crystal structure of the N-domains in complex with lactose

The three-dimensional structure of the N-domain of wild type hGal-8 loaded with lactose was determined at 1.35 Å resolution. The final model comprised 148 protein residues, 252 water molecules, one lactose molecule, one Zn^{2+}, one Na^{+} and one molecule of glycerol (for details, please see Table 2). The structure presents the canonical "jelly roll" topology, with two sets of antiparallel β-sheets forming the β-sandwich arrangement typical of all known mammalian galectins. In addition to the common six- and five-stranded β-sheets (S1-S6, F1-F5), the structure reveals the presence of a previously unassigned extra β-strand [43, 44], termed F0 (Fig. 3A), as similarly found for the separate N-domain of human galectin-9 [45]. Strand F0 runs in antiparallel manner to the carboxy-terminal F1 strand, the two strands establishing a network of backbone hydrogen bonds (Fig. 3B). The carbohydrate-binding site is formed by residues from β-strands S4-S6 (Fig. 3C). Arg-69 interacts through hydrogen bonds with the OH-3 of the glucose unit of lactose, while the galactose unit stacks with the indole ring of Trp-86 as well as establishes direct hydrogen bonds with Arg-45, His-65, Asn-67, Asn-79 and Glu-89 and, through a water-mediated network, with Arg-59 and Trp-86. A strong electron density close to OH-3' of galactose was modeled as a Na^{+} cation. In addition, a molecule of glycerol (the additive used for cryoprotection) was identified close to Tyr-141.

Superposition with the structure available for the same complex at lower resolution (2.33 Å, PDB entry 3AP4 [43]) showed only minor differences (Fig. 3D), as reflected by an RMSD value between Cα atoms of 0.23 Ǻ, whereas comparison with the structure of the ligand-free form (PDB entry 3AP5, RMSD value 0.33 Ǻ for all Cα atoms) revealed changes in the position of residues at the binding site. Arg-69 and Glu-89 move towards lactose for establishing hydrogen bonds with the sugar. Furthermore, the position of Arg-45, which exhibits a double conformation in the ligand-free form, is found to be stabilized upon lactose binding, and a small shift (around 1 Å) of Trp-86 is also noticed. Yet, the biggest changes are observed for the loop connecting strands S3 and S4, which moves 1.5 Å towards the sugar, thereby narrowing the ligand-binding site and displacing 2 Å the side chain of Arg-59 from its position in the ligand-free form (Fig. 3D). All these ligand-induced rearrangements within the carbohydrate-binding site are likely to affect the environment of Trp-86 and Tyr-141, hereby providing a structural explanation for the differences observed in the near-UV CD spectra of the hGal-8 N-domain in the absence and presence of lactose. Of note, no similar changes are detected when comparing the ligand-free and lactose-loaded crystal structures available for human galectins-1, -3, -7 and -9, the readjustment observed for the hGal-8 N-domain thus appearing as a distinctive feature of this CRD.

The crystal structure of the mutant N-domain in complex with lactose was solved in parallel at 1.9 Å resolution. A total of 291 residues (145 in chain A and 146 in chain B), 2 molecules of lactose and 99 of

water constituted the final model (Table 2). In the crystallographic dimer, the long loop connecting strands F5 and S2 of one subunit establishes 9 hydrogen bonds and 4 salt bridges with the edge of strands F2-F3-F4 from the neighbor subunit. According to the complex formation significance score of the PISA web server [51], these interactions, however, may not be sufficient for physiological homodimer formation, in agreement with the monomeric status determined in solution by gel filtration chromatography and analytical ultracentrifugation.

A superposition of the structures of the F19Y and wild type N-domains reflected a highly similar global structure (Fig. 4A), with RMSD between Cα atoms being only 0.2 Å. Furthermore, the carbohydrate-binding sites exhibit identical architecture and geometry of lactose binding (Fig. 4B). Focusing on the position 19, the lateral chain of Tyr-19 superposes with wild type Phe-19 (Fig. 4C). However, the presence of the tyrosine hydroxyl group induces a shift of β-strand F0, leading to a series of displacements of residues from position 15 towards the N-terminus, with Ile-11, Tyr-13, Asn-14 and Pro-15 being significantly displaced from the position occupied in the wild type structure. In the full-length protein, this strand precedes the linker peptide that connects the N- and C-terminal domains. Previous studies on the impact on human galectin-1 of single-site mutations introduced at some distance from the binding site revealed that the effects are not locally confined [6]. On the contrary, they are transmitted throughout the CRD even affecting the binding thermodynamics. Thus, it seems likely that the shift of the F0 strand observed in the F19Y variant may also be transmitted to other sections of the protein, even beyond the N-domain, thereby accounting for the observed differences in thermal stability. This reasoning prompted us to examine whether the thermodynamic parameters of ligand binding to the N-domains and to the full-length proteins could also be affected by the F19Y substitution. Thus, we next performed isothermal titration calorimetry (ITC), testing lactose and 3'-sialyllactose as in the CD-based experiments.

Thermodynamic parameters of ligand binding

Representative curves of calorimetric titrations with lactose and also with 3'-sialyllactose are shown in Fig. 5, and the thermodynamic parameters derived from the analysis are summarized in Table 3. Titration curves of full-length proteins with lactose were fitted to two-independent sets of sites (one per CRD) with binding constants of 1.1×10^{4} M^{-1} and 310-460 M^{-1}. Comparison of binding parameters with those obtained for lactose binding to the N-domains enabled the unambiguous assignment of the high-affinity site to the N-terminal CRD (Table 3). By contrary, data obtained upon titration of full-length proteins with 3'-sialyllactose were consistent with the presence of only one set of binding sites per molecule; the one of the N-terminal CRD as titration of separate domains with 3'-sialyllactose proved (Fig. 5 and Table 3). Of note, the thermodynamic parameters indicate differences in ΔH and ΔS contributions to the affinity in a compensatory manner, with more negative values in the case of the full-length mutant and lactose (high-

affinity site). In all other cases both contributions are more negative for the wild type complexes.

In spite of sharing almost identical binding-site architectures, the enthalpic contribution to the binding of lactose to the separate F19Y N-domain was about 1 kcal/mol lower than for the wild type domain. This smaller ΔH was counterbalanced by a reduced entropic penalty, resulting in same binding affinities. It seems therefore reasonable to assume that observed differences in ΔH and ΔS can arise from distinct contributions of ligand-induced protein rearrangements beyond the carbohydrate-binding site. On the other hand, the affinity of both N-terminal sites for 3'-sialyllactose was comparable and one order of magnitude higher than its affinity for lactose, generally due to a substantially greater enthalpic contribution to the binding (Table 3). This increase is most probably attributable to near contacts of the sialic acid moiety with the protein, as observed in the crystal structure of the wild type N-domain – 3'-sialyllactose complex [43]. In direct comparison, the enthalpic contribution to the binding of 3'-sialyllactose to the F19Y N-domain is about 0.3 kcal/mol smaller than for the wild type domain, again being counterbalanced by a slightly lower entropy loss. The difference between binding thermodynamics is notably more pronounced for the full-length proteins than for the respective separate N-domains (Table 3). With ΔΔH of -1.3 kcal/mol and ΔΔS at -4.2 cal/(mol·K) and standard deviations of +/-(0.1/0.1) kcal/mol or +/-(0.2/0.7) cal/(mol·K), the F19Y substitution makes its presence clearly felt in the full-length proteins.

The lower accuracy of the thermodynamic parameters derived for lactose binding to the C-terminal CRD (low-affinity site) in full-length proteins precludes to reliably postulate a small effect of F19Y substitution at the site distant to position 19 (Table 3). The listed indications for alterations in binding parameters (but not overall affinity) of the N-domains and the full-length proteins for two types of free ligand prompted a comparative evaluation of the cell binding and effector activities of the wild type and F19Y hGal-8 proteins.

Comparison of functional properties

Since galectins are known to affect cell properties after carbohydrate-dependent surface binding, we measured affinity of association to cells and effects of this binding on cell growth *in vitro* for the two proteins in direct comparison. As test systems, we selected a line already studied with several other galectins (neuroblastoma), a leukemic and four carcinoma lines, thus representing different histogenetic origins. Binding to human neuroblastoma (SK-N-MC) cells was nearly completely blocked by the presence of glycan inhibitors (150 mM lactose and 0.5 mg/ml asialofetuin) in both cases. Measuring the extent of binding with increasing lectin concentrations, the two curves for the wild type protein and its variant were very similar, with calculated K_D-values (B_{max}-values) of about 118 nM (4.6×10^5 molecules/cell) for the wild type protein and 102 nM (4.9×10^5 molecules/cell) for the variant (Fig. 6).

The binding behavior at the level of cells therefore was in line with the similar binding affinities determined for simple sugars by ITC. Of note, the presence of ganglioside GM1-specific cholera toxin B-subunit only slightly reduces the extent of cell binding (Fig. 6). In contrast to homodimeric galectins-1, -2 and -7 whose sites of binding are masked by the bacterial protein, ganglioside GM1 apparently is not a key counterreceptor of hGal-8 on these cells. However, this difference in target specificity did not abolish the activity of hGal-8 as growth inhibitor. Both proteins were significantly active in this respect (Fig. 7). Besides presenting the effects of each protein relative to the control without galectin (Fig. 7A), we set the data obtained with the two proteins also into direct comparison. This data processing came up with a slightly increased inhibitory activity of the F19Y variant at $p \leq 0.05$ (Fig. 7B). To further examine this feature comparatively with cells of different histogenetic origin we next tested erythroleukemia (K562) cells.

Again, the two proteins reduced cell growth significantly, and the variant clearly proved to be more potent in this respect (Fig. 8). At the highest concentration tested, the significance threshold of $p \leq 0.005$ was surpassed. Thus, the detected, seemingly subtle structural changes arising from the F19Y substitution do appear to have functional implication. As documented above, extent of erythroleukemia cell binding appeared to be rather similar, as was sensitivity to the presence of lactose (Fig. 9). A similar pattern for affecting cell proliferation was determined on four human colon adenocarcinoma lines (Fig. 10). In these cases, the SNP-derived variant form was more potent to reduce growth, with differences in concentration dependence between the lines. Overall, the data on proliferation support to the presence of a quantitative difference between the two types of proteins in growth regulation of the tested cell systems.

Besides characterizing the effect on proliferation, likely caused by cross-linking counterreceptors in *cis*-configuration on the cell surface, we determined the potency of the proteins to form cell aggregates in classical haemagglutination. Herein, rabbit erythrocytes are cross-linked in *trans*-configuration by the bivalent proteins. The minimal concentration for a positive response was 600 ng/assay for the wild type protein, and 1000-1200 ng/assay for the variant. Sensitivity for lactose presence was higher for the variant (concentrations above 0.8 mM blocked agglutination at 1400 ng) than for the wild type protein (at least 1.5 mM required at 1000 ng).

Discussion

The natural occurrence of SNP-based protein variants presents an attractive model for analyzing structure-activity relationships. Whether the amino acid substitution can affect protein properties will have a notable bearing on trait prediction from genotype data, prone to diverse pitfalls [52]. We have here initiated this line of research for galectins, focusing on the F19Y substitution in the N-terminal CRD of hGal-8. The seemingly small change, i.e. addition of a hydroxyl group to an aromatic side chain, is not at a site of direct contact with the carbohydrate ligand but part of the F0 β-strand. Our biophysical

characterization combining the analysis of the crystallized protein and protein in solution teaches a series of lessons:

a.) All the proteins are present as monomers in solution, tested under conditions that have detected concentration-dependent changes in the status of aggregation of a plant lectin [53], and the hydrodynamic behavior is not measurably altered by the Phe to Tyr replacement.

b.) The substitution causes displacements of amino acid side chains in its vicinity, defined by crystallography.

c.) The consequences of the F19Y substitution are not locally confined. Instead, they apparently extend through the tandem-repeat-type protein, as revealed by measuring thermal denaturation and thermodynamic binding parameters. Such an effect might be facilitated by the shift of the F0 β-strand caused by the F19Y substitution, as this strand positionally precedes the linker connecting the N- and C-terminal CRDs.

d.) The characteristics of impact on ligand binding, in comparison to the wild type protein, appear to depend on the nature of the carbohydrate ligand. The measurements with two common ligands (lactose and its 3'-sialylated derivative) led to distinct data sets in terms of enthalpy/entropy compensation, suggesting that the same may likely be true for cellular activities, if involving different counterreceptors. Of note, such a case has recently been reported for an engineered mutant of human galectin-2 [54].

At this stage, the structural analysis of the variant allowed to infer that the single-site substitution has consequences beyond its local vicinity, in principle in line with the hypothesis of possible functional correlations. On the level of the tested cell types, binding properties were rather similar for the wild type and variant proteins. As seen in ITC, affinity was not affected, when quantitating extent of binding of labeled (iodinated or biotinylated) protein. Of note, the application of the cholera toxin B-subunit revealed that ganglioside GM1 is not a neuroblastoma cell counterreceptor for galectin-8, as it is for homodimeric galectins [55-57] and also for tandem-repeat-type galectin-4 [58]. As consequence, galectin-8 will then also likely not trigger ganglioside GM1-dependent anergy/apoptosis of activated effector T cells, as galectin-1 does [59, 60]. Obviously, this observation signifies that members of the galectin family can differ drastically in distinct aspects of target selection.

Probing into functional aspects, the variant proved to be more active as growth inhibitor on the three types of tumor cells than the wild type protein. Structural changes induced by the presence of the hydroxyl group thus translate into enhanced effector potency, with quantitative differences between lines of the same histogenetic origin. In addition to measuring proliferation, we determined the two proteins´capacity for cross-linking cells, using trypsin-treated, glutaraldehyde-fixed rabbit erythrocytes. Stable aggregate formation (with interactions established in *trans*) required less protein and was more resistant to presence of lactose with the wild type protein. These results on forming stable aggregates

point to the perspective to further examine binding properties when applying force. Equally important, since hGal-8 is known to be present in cytoplasm and nuclei (21-26), it should be kept in mind that not only glycans are physiologic counterreceptors. Galectins in general are multifunctional, and are endowed with capacity to interact with distinct peptide motifs besides their reactivity with glycans, and this at different sites [61]. A protein interacting with hGal-8´s C-domain is an autophagy cargo receptor [46, 47, 62]. In this sense, our results on actual effects of a subtle change encourage further detailed comparisons with the two proteins using various binding partners. Along this line, the other known polymorphisms of the hGal-8 gene clearly deserve attention, as do the aspects of physiological linker-length variation by alternative splicing and of sequence divergence in phylogenesis.

Beyond the given focus on (ga)lectins, our results strongly recommend the analysis in other cases beyond evident loss-of-function occurrence, as in dectin-1´s early-stop-codon (Y238X) mutant or in variants in which sites with direct ligand contact are hit, as seen for K131Q in siglec-9 [63, 64]. In addition to galectins, ficolin-L (product of the FCN2 gene), here variants exhibiting substitutions T236M/A258S in the vicinity of the contact site [65-67], or the product of the S128R polymorphism in E-selectin, with the change located in the epidermal growth factor-like domain and triggering tethering of myeloid cells [68], are other examples of candidates for respective analysis along the lines presented herein.

Considering the aspect of interplay of (ga)lectins with specific counterreceptors inspires the idea of a pairing in association studies on SNPs. In fact, functional correlations to glycogenes, via orchestrated co-regulation by a master regulator such as a tumor suppressor [69-71] or the detection of importance of distinct glycan determinants in tumor growth/host defense by Gal-1 [72, 73], can identify partner genes for setting up meaningful combinations. Case studies on core 2 N-acetylglucosaminyltransferase connect a variant (i.e. S158C) to Gal-1-induced T lymphoma growth inhibition [74] and a second variant (i.e. V152I) to increased risk of prostate cancer [75]. Collectively, our data thus reveal an unsuspected broadness of effects of the natural deviation at a single site (F19Y), establishing a precedent for further investigations, and reported information points to the potential of guided candidate selection to design paired association studies based on functional cooperation.

Experimental procedures

Protein preparation

Wild type and variant proteins (full-length proteins with 34-amino-acid linker and the separate N-domains) were obtained by recombinant production followed by chromatographic purification and rigorous quality controls as described [18, 41]. Protein concentrations were routinely determined using the theoretical molar adsorption coefficients.

Mass spectrometry

MALDI-MS measurements were performed on a Voyager DE-PRO mass spectrometer (Applied Biosystems, Foster City, CA) equipped with a pulsed nitrogen laser (λ = 337 nm, 3 ns pulse width, and 3 Hz frequency) and a delayed extraction ion source. Ions generated by laser desorption of samples were introduced into a time of flight analyzer (1.3 m flight path) with an acceleration voltage of 25 kV, 90-93% grid voltage, 0.01-0.1% ion guide wire voltage, and a delay time of 350 ns in the linear positive ion mode. Mass spectra of the separate N-domains were obtained over the m/z range 10-25 ku, using thioredoxin (*Escherichia coli*, oxidized) and apomyoglobin (Calibration Mixture 3 of Sequazyme Peptide Mass Standards Kit; Applied Biosystems) for external mass calibration. Spectra of the full-length proteins were obtained over the m/z range 25-50 ku. Carbonic anhydrase and enolase from *Saccharomyces cerevisiase* were used in this case for external mass calibration. Sinapinic acid at 10 mg/mL in TFA 0.3%:acetonitrile 70:30 (vol/vol) was used as matrix. Samples were mixed with the matrix at a ratio of ~1:4 (vol/vol), and 1 µl of this solution was spotted onto a flat stainless-steel sample plate and dried in air.

Analytical gel filtration and ultracentrifugation

Gel filtration was performed with a Superose 12 10/300 GL column (void volume: 7.8 ml; GE Healthcare) at a flow rate of 0.5 ml/min in 5 mM sodium phosphate buffer (pH 7.2) containing 0.2 M NaCl (PBS), 4 mM β-mercaptoethanol, 0.02% NaN_3, and 0.1 M lactose. BSA (66 kDa), ovalbumin (42.7 kDa), carbonic anhydrase (29 kDa), cytochrome C (12.4 kDa), aprotinin (6.5 kDa), and vitamin B_{12} (1.35 kDa) were chromatographed under similar conditions for column calibration. Sedimentation equilibrium experiments were carried out by centrifugation of 80 µl samples, adjusted to different protein concentrations, at 15,200, 18,000 and 25,000 rev/min for the N-domains and 10,000, 11,800 and 20,400 rev/min for the full-length proteins, as described previously [53]. Sedimentation velocity experiments were run at 45,000 rev/min for 5 h using 400 µl samples. All measurements were performed at 20 ºC in an Optima XL-A analytical ultracentrifuge (Beckman Coulter, Krefeld, Germany) with an AN50-Ti rotor. Weight-average molecular weights were calculated using the HeteroAnalysis program, version 1.1.2 (http://www.biotech.uconn.edu/auf/). Differential sedimentation coefficients were calculated by least-squares boundary modeling of the experimental data with the program SEDFIT [76]. Solvent density and viscosity at 20 °C were computed using the Sednterp software. The partial specific volume and degree of hydration of the proteins were calculated from the amino acid composition using the same program.

Circular dichroism

CD spectra were measured with a J-810 spectropolarimeter, equipped with a Peltier temperature control system, using a bandwidth of 1 nm and a response time of 4 s. Far-UV spectra were recorded in 0.1 cm

path-length quartz cells at a protein concentration of 0.2 mg/ml, and near-UV spectra at 1.0 mg/ml in 1 cm path-length cuvettes. Routinely, the corresponding buffer baseline was subtracted. The spectra were representative of three to four independent measurements. Thermal denaturation experiments were carried out by increasing the temperature from 20 °C to 90 °C at a scanning rate of 0.66 °C/min. Variations in ellipticity at 218 nm were monitored at steps of 0.2 °C. Thermal denaturation profiles were described in terms of the following sigmoidal function:

$$\Theta(T) = \Theta_D(T) - [\Theta_D(T) + \Theta_N(T)] / \{1+\exp[A(T-T_{1/2})/RTT_{1/2}]\}$$

where T is the absolute temperature, $T_{1/2}$ is the half-transition temperature, R is the gas constant, A is the constant accounting for the variation with temperature of native and denatured states, and $\Theta_D(T)$ and $\Theta_N(T)$ denote the ellipticity of the denatured and native states at temperature T. Θ_D and Θ_N were approximated as linear functions of temperature, as described previously [77]. The fitting of the sigmoid to the experimental data was carried out by non-linear minimum-squares, using the Origin software.

Crystallization, data collection and processing

The N-domains of wild type and F19Y proteins were concentrated to 20 mg/ml using Amicon Ultra centrifugal filter units (Millipore). Crystals were grown using the sitting drop vapor diffusion method at 22 ºC, by mixing 1 µl of protein solution with 1 µl of reservoir solution. The wild type N-domain was crystallized in 20% w/v PEG 4000, 100 mM HEPES sodium salt (pH 7.2) and 10 mM zinc chloride. The N-domain of the F19Y mutant was crystallized in 30% w/v PEG 4000, 100 mM MES sodium salt (pH 6.5). Crystals were transferred to a cryoprotectant solution consisting in the reservoir solution supplemented with 20% v/v glycerol (wild type N-domain) or 20% ethylene glycol (F19Y mutant N-domain) and subsequently flash-cooled in liquid nitrogen.

X-Ray diffraction data were collected at Beamlines ID23-1 and ID29 of the European Synchrotron Radiation Facility (ESRF) and at the PROXIMA-1 beamline SOLEIL Synchrotron (France). Crystallographic data were processed using XDS [78] and Aimless [79]. The space group of crystals and unit cell dimensions are specified in Table 2. The structures have been solved in two different crystal forms. For the wild type N-domain, the tetragonal crystals contained one monomer, while the orthorhombic crystals of the F19Y variant contained a dimer in the asymmetric unit. Likewise, previously reported crystal structures of the wild type N-domain in complex with several ligands were solved in three crystal forms with dissimilar number of protein molecules (from 1 to 4) in the asymmetric unit [43], the different crystal packing interactions not affecting the overall fold. The Mathews coefficient for a monomer in the asymmetric unit was 2.74 Å^3/Da for the wild type N-domain and 2.15 Å^3/Da for a dimer of the F19Y mutant, which correspond to a solvent content of 55.2% and 42.7%, respectively [80].

The atomic coordinates and structure factors (codes 4BMB, 4BME) have been deposited in the Protein

Data Bank.

Structure determination and refinement

The structures were determined with the molecular replacement method implemented in the Phenix software suite [81], using the structure at 1.92 Å-resolution available for the N-domain of wild type hGal-8 (Protein Data Bank entry code 2YV8) as model. The translation-libration-screw (TLS) groups were defined using the TLMSD server [82]. Repetitive cycles of refinement and manual building were done using Phenix [81] and Coot [83]. Water molecules were added to the model during the last steps of refinement. Final R/R_{free} values were 0.16/0.18 and 0.23/0.28 for the wild type and F19Y mutant structures, respectively. Validation of final models was performed using the Molprobity software [84]. Protein-protein interactions were analyzed using the PISA web server [51]. Details of structure refinement and quality of final models are presented in Table 2. Structural figures were generated using Pymol [85].

Isothermal titration calorimetry

Calorimetric titrations were performed at 25 ºC with a Microcal VP-ITC microcalorimeter (GE Healthcare, Northampton, MA, USA), as recently described for chicken galectins [86]. The proteins were exhaustively dialyzed against PBS containing 4 mM β-mercaptoethanol (PBS_{β}) and sugar solutions were prepared with the final dialysate. Titrations were performed by stepwise injections of the sugar-containing solution into the reaction cell loaded with the protein solution at concentrations of 165-170 μM. The heat of ligand dilution was determined separately and subtracted from the total heat produced following each injection. Titration data were analyzed assuming binding models with either one or two classes of binding sites using the Microcal-ITC Origin software.

Cell assays

Haemagglutination was performed with trypsin-treated glutaraldehyde-fixed rabbit erythrocytes in V-shaped microtiter plates in V-shaped microtiter plates using 25 μL of a 5% suspension and 25 lL galectin-containing solution (serial two-fold dilution of freshly prepared stock solutions of different initial concentrations). Cells for the proliferation assays were grown in RPMI 1640 medium supplemented with 1.75% fetal calf serum (erythroleukemia (K562) cells), Eagle's minimal essential medium containing 10% fetal calf serum (neuroblastoma (SK-N-MC) cells) and Dulbecco's minimal essential medium containing 10% fetal calf serum (colon adenocarcinoma cell lines derived from microsatellite-instable (HCT116, Colo60H) or microsatellite-stable (SW480, SW707) tumors). All media additionally contained antibiotics (100 U/ml penicillin and 100 μg/ml streptomycin). Lectin binding to cell surfaces was analyzed either with biotinylated proteins in FACScan analysis using fluorescent streptavidin-R-phycoerythrin as reporter

or with iodinated proteins (wild type protein: 195 KBq/µg, F19Y variant: 219 KBq/µg, prepared with carrier-free $Na^{125}I$ (Hartmann Analytic, Braunschweig, Germany) and Iodobeads (Pierce, Bonn, Germany)) following optimized protocols [55, 56]. Competition assays with cholera toxin B-subunit (Sigma, Munich, Germany) were performed with iodinated galectins, adding the ganglioside GM1-specific probe 1 h prior to the labeled protein as described [58]. Cell growth was assessed in parallel assays using the blue chromogen 3-(4,5-dimethyl-thiazol-3-yl)-2,5diphenyl-tetrazolium bromide (0.5 mg/ml; Sigma) or a commercial kit (CellTiter 96; Promega, Mannheim, Germany) [58].

Acknowledgments

This work was generously supported by grants BFU2009-10052, BFU2011-24615, BFU2012-36825 and CSD2009-00088 from the Spanish Ministry of Science and Innovation, the CIBER of Respiratory Diseases (CIBERES), an initiative from ISCIII, the Regional Government of Madrid (S2010/BMD-2353 and S2010/BMD-2457) and funding from the European Union's Seventh Framework Program FP7/2007-2013 under REA grant agreements no. 260600 ("GlycoHIT") and 317297 ("GLYCOPHARM"), F.M.R. by an ESF/CSIC funded JAE-Doc Contract. We also cordially thank the staffs of beamlines ID23-1 and ID29 at ESRF (Grenoble) and Proxima I (SOLEIL) as well as Dr. B. Friday for their invaluable help.

References

1. Gabius H-J. ed (2009) *The Sugar Code. Fundamentals of glycosciences.* Wiley-VCH, Weinheim, Germany.
2. Gabius H-J, André S, Jiménez-Barbero J, Romero A & Solís D (2011) From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem Sci* **36**, 298-313.
3. Liu F-T, Yang R-Y & Hsu DK (2012) Galectins in acute and chronic inflammation. *Ann N Y Acad Sci* **1253**, 80-91.
4. Kaltner H & Gabius H-J (2012) A toolbox of lectins for translating the sugar code: the galectin network in phylogenesis and tumors. *Histol Histopathol* **27**, 397-416.
5. Hirabayashi J & Kasai K-i (1991) Effect of amino acid substitution by sited-directed mutagenesis on the carbohydrate recognition and stability of human 14-kDa β-galactoside-binding lectin. *J Biol Chem* **266**, 23648-2353.
6. López-Lucendo MF, Solís D, André S, Hirabayashi J, Kasai K-i, Kaltner H, Gabius H-J & Romero A (2004) Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. *J Mol Biol* **343**, 957-970.

7. Ozaki K, Inoue K, Sato H, Iida A, Ohnishi Y, Sekine A, Sato H, Odashiro K, Nobuyoshi M, Hori M, Nakamura Y & Tanaka T (2004) Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-α secretion in vitro. *Nature* **429**, 72-75.
8. Iida A, Ozaki K, Tanaka T & Nakamura Y (2005) Fine-scale SNP map of an 11-kb genomic region at 22q13.1 containing the galectin-1 gene. *J Hum Genet* **50**, 42-45.
9. Li W, Xu J, Wang X, Chen J, Zhang C, Sun K & Hui R (2010) Lack of association between lymphotoxin-α, galectin-2 polymorphisms and coronary artery disease: a meta-analysis. *Atherosclerosis* **208**, 433-436.
10. Pál Z, Antal P, Millinghoffer A, Hullám G, Pálóczi K, Tóth S, Gabius H-J, Molnár MJ, Falus A & Buzás EI (2010) A novel galectin-1 and interleukin 2 receptor β haplotype is associated with autoimmune myasthenia gravis. *J Neuroimmunol* **229**, 107-111.
11. Lotz MM, Andrews CW Jr, Korzelius CA, Lee EC, Steele GD Jr, Clarke A & Mercurio AM (1993) Decreased expression of Mac-2 (carbohydrate binding protein 35) and loss of its nuclear localization are associated with the neoplastic progression of colon carcinoma. *Proc Natl Acad Sci USA* **90**, 3466-3470.
12. Balan V, Nangia-Makker P, Schwartz AG, Jung YS, Tait L, Hogan V, Raz T, Wang Y, Yang ZQ, Wu GS, Guo Y, Li H, Abrams J, Couch FJ, Lingle WL, Lloyd RV, Ethierb SP, Tainsky MA & Raz A (2008) Racial disparity in breast cancer and functional germ line mutation in galectin-3 (rs4644): a pilot study. *Cancer Res* **68**, 10045-10050.
13. Mazurek N, Byrd JC, Sun Y, Ueno S & Bresalier RS (2011) A galectin-3 sequence polymorphism confers TRAIL sensitivity to human breast cancer cells. *Cancer* **117**, 4375-4380.
14. Meyer A, Coinac I, Bogdanova N, Dubrowinskaja N, Turmanov N, Haubold S, Schürmann P, Imkamp F, von Klot C, Merseburger AS, Machtens S, Bremer M, Hillemanns P, Kuczyk MA, Karstens JH, Serth J & Dörk T (2013) Apoptosis gene polymorphisms and risk of prostate cancer: a hospital-based study of German patients treated with brachytherapy. *Urol Oncol* **31**, 74-81.
15. Hu CY, Chang SK, Wu CS, Tsai WI & Hsu PN (2011) Galectin-3 gene (LGALS3) +292C allele is a genetic predisposition factor for rheumatoid arthritis in Taiwan. *Clin Rheumatol* **30**, 1227-1233.
16. Chen HJ, Zheng ZC, Yuan BQ, Liu Z, Jing J & Wang SS (2012) The effect of galectin-3 genetic variants on the susceptibility and prognosis of gliomas in a Chinese population. *Neurosci Lett* **518**, 1-4.
17. Wu F, Hu N, Li Y, Bian B, Xu G & Zheng Y (2012) Galectin-3 genetic variants are associated with platinum-based chemotherapy response and prognosis in patients with NSCLC. *Cell Oncol* **35**, 175-180.

18. Pál Z, Antal P, Srivastava SK, Hullám G, Semsei AF, Gál J, Svébis M, Soós G, Szalai C, André S, Gordeeva E, Nagy G, Kaltner H, Bovin NV, Molnár MJ, Falus A, Gabius H-J & Buzás EI (2012) Non-synonymous single nucleotide polymorphisms in genes for immunoregulatory galectins: association of galectin-8 (F19Y) occurrence with autoimmune diseases in a Caucasian population. *Biochim Biophys Acta* **1820**, 1512-1518.
19. Hadari YR, Eisenstein M, Zakut R & Zick Y (1997) Galectin-8: on the road from structure to function. *Trends Glycosci Glycotechnol* **9**, 103-112.
20. Kaltner H, Solís D, André S, Lensch M, Manning JC, Mürnseer M, Sáiz JL & Gabius H-J (2009) Unique chicken tandem-repeat-type galectin: implications of alternative splicing and a distinct expression profile compared to those of the three proto-type proteins. *Biochemistry* **48**, 4403-44016.
21. Danguy A, Rorive S, Decaestecker C, Bronckart Y, Kaltner H, Hadari YR, Goren R, Zick Y, Petein M, Salmon I, Gabius H-J & Kiss R (2001) Immunohistochemical profile of galectin-8 expression in benign and malignant tumors of epithelial, mesenchymatous and adipous origins, and of the nervous system. *Histol Histopathol* **16**, 861-868.
22. Nagy N, Bronckart Y, Camby I, Legendre H, Lahm H, Kaltner H, Hadari Y, van Ham P, Yeaton P, Pector JC, Zick Y, Salmon I, Danguy A, Kiss R & Gabius H-J (2002) Galectin-8 expression decreases in cancer compared with normal and dysplastic human colon tissue and acts significantly on human colon cancer cell migration as a suppressor. *Gut* **50**, 392-401.
23. Langbein S, Brade J, Badawi JK, Hatzinger M, Kaltner H, Lensch M, Specht K, André S, Brinck U, Alken P & Gabius H-J (2007) Gene-expression signature of adhesion/growth-regulatory tissue lectins (galectins) in transitional cell cancer and its prognostic relevance. *Histopathology* **51**, 681-690.
24. Cludts S, Decaestecker C, Mahillon V, Chevalier D, Kaltner H, André S, Remmelink M, Leroy X, Gabius H-J & Saussez S (2009) Galectin-8 up-regulation during hypopharyngeal and laryngeal tumor progression and comparison with galectins-1, -3 and -7. *Anticancer Res* **29**, 4933-4940.
25. Saussez S, de Leval L, Decaestecker C, Sirtaine N, Cludts S, Duray A, Chevalier D, André S, Gabius H-J, Remmelink M & Leroy X (2010) Galectin fingerprinting in Warthin's tumors: lectin-based approach to trace its origin? *Histol Histopathol* **25**, 541-550.
26. Remmelink M, de Leval L, Decaestecker C, Duray A, Crompot E, Sirtaine N, André S, Kaltner H, Leroy X, Gabius H-J & Saussez S (2011) Quantitative immunohistochemical fingerprinting of adhesion/growth-regulatory galectins in salivary gland tumours: divergent profiles with diagnostic potential. *Histopathology* **58**, 543-556.
27. Hadari YR, Arbel-Goren R, Levy Y, Amsterdam A, Alon R, Zakut R & Zick Y (2000) Galectin-8 binding to integrins inhibits cell adhesion and induces apoptosis. *J Cell Sci* **113**, 2385-2397.

28. Levy Y, Arbel-Goren R, Hadari YR, Eshhar S, Ronen D, Elhanany E, Geiger B & Zick Y (2001) Galectin-8 functions as a matricellular modulator of cell adhesion. *J Biol Chem* **276**, 31285-31295.
29. Arbel-Goren R, Levy Y, Ronen D & Zick Y (2005) Cyclin-dependent kinase inhibitors and JNK act as molecular switches, regulating the choice between growth arrest and apoptosis induced by galectin-8. *J Biol Chem* **280**, 19105-19114.
30. Sebban LE, Ronen D, Levartovsky D, Elkayam O, Caspi D, Aamar S, Amital H, Rubinow A, Golan I, Naor D, Zick Y & Golan I (2007) The involvement of CD44 and its novel ligand galectin-8 in apoptotic regulation of autoimmune inflammation. *J Immunol* **179**, 1225-1235.
31. Tribulatti MV, Mucci J, Cattaneo V, Agüero F, Gilmartin T, Head SR & Campetella O (2007) Galectin-8 induces apoptosis in the $CD4^{high}CD8^{high}$ thymocyte subpopulation. *Glycobiology* **17**, 1404-1412.
32. Tribulatti MV, Cattaneo V, Hellman U, Mucci J & Campetella O (2009) Galectin-8 provides costimulatory and proliferative signals to T lymphocytes. *J Leukoc Biol* **86**, 371-380.
33. Cattaneo V, Tribulatti MV & Campetella O (2011) Galectin-8 tandem-repeat structure is essential for T-cell proliferation but not for co-stimulation. *Biochem J* **434**, 153-160.
34. Tsai CM, Guan CH, Hsieh HW, Hsu TL, Tu Z, Wu KJ, Lin CH & Lin KI (2011) Galectin-1 and galectin-8 have redundant roles in promoting plasma cell formation. *J Immunol* **187**, 1643-1652.
35. Massardo L, Metz C, Pardo E, Mezzano V, Babul M, Jarpa E, Guzmán AM, André S, Kaltner H, Gabius H-J, Jacobelli S, González A & Soza A (2009) Autoantibodies against galectin-8: their specificity, association with lymphopenia in systemic lupus erythematosus and detection in rheumatoid arthritis and acute inflammation. *Lupus* **18**, 539-546.
36. Sarter K, Janko C, André S, Muñoz LE, Schorn C, Winkler S, Rech J, Kaltner H, Lorenz HM, Schiller M, Andreoli L, Manfredi AA, Isenberg DA, Schett G, Herrmann M & Gabius H-J (2013) Autoantibodies against galectins are associated with antiphospholipid syndrome in patients with systemic lupus erythematosus. *Glycobiology* **23**, 12-22.
37. Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, Hirashima M, Urashima T, Oka T, Futai M, Müller WEG, Yagi F & Kasai K-i (2002) Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim Biophys Acta* **1572**, 232-254.
38. Ideo H, Seko A, Ishizuka I & Yamashita K (2003) The N-terminal carbohydrate recognition domain of galectin-8 recognizes specific glycosphingolipids with high affinity. *Glycobiology* **13**, 713-723.
39. Carlsson S, Oberg CT, Carlsson MC, Sundin A, Nilsson UJ, Smith D, Cummings RD, Almkvist J, Karlsson A & Leffler H (2007) Affinity of galectin-8 and its carbohydrate recognition domains for ligands in solution and at the cell surface. *Glycobiology* **17**, 663-676.

40. Stowell SR, Arthur CM, Slanina KA, Horton JR, Smith DF & Cummings RD (2008) Dimeric galectin-8 induces phosphatidylserine exposure in leukocytes through polylactosamine recognition by the C-terminal domain. *J Biol Chem* **283**, 20547-2059.
41. Vokhmyanina OA, Rapoport EM, Ryzhov IM, Korchagina EY, Pazynina GV, Severov VV, Kaltner H, André S, Gabius H-J & Bovin NV (2011) Carbohydrate specificity of chicken and human tandem-repeat-type galectins-8 in composition of cells. *Biochemistry (Mosc)* **76**, 1185-1192.
42. Vokhmyanina OA, Rapoport EM, André S, Severov VV, Ryzhov I, Pazynina GV, Korchagina E, Gabius H-J & Bovin NV (2012) Comparative study of the glycan specificities of cell-bound human tandem-repeat-type galectins-4, -8 and -9. *Glycobiology* **22**, 1207-1217.
43. Ideo H, Matsuzaka T, Nonaka T, Seko A & Yamashita K (2011) Galectin-8 N-domain recognition mechanism for sialylated and sulfated glycans. *J Biol Chem* **286**, 11346-11355.
44. Yoshida H, Yamashita S, Teraoka M, Itoh A, Nakakita S, Nishi N & Kamitori S (2012) X-Ray structure of a protease-resistant mutant form of human galectin-8 with two carbohydrate recognition domains. *FEBS J* **279**, 3937-3951.
45. Solís D, Maté MJ, Lohr M, Ribeiro JP, López-Merino L, André S, Buzamet E, Cañada FJ, Kaltner H, Lensch M, Ruiz FM, Haroske G, Wollina U, Kloor M, Kopitz J, Sáiz JL, Menéndez M, Jiménez-Barbero J, Romero A & Gabius H-J (2010) N-Domain of human adhesion/growth-regulatory galectin-9: preference for distinct conformers and non-sialylated N-glycans and detection of ligand-induced structural changes in crystal and solution. *Int J Biochem Cell Biol* **42**, 1019-1029.
46. Kim B-W, Hong SB, Kim JH, Kwon DH & Song HK (2013) Structural basis for recognition of autophagic receptor NDP52 by the sugar receptor galectin-8. *Nat Commun* **4**, 1613.
47. Li S, Wandel MP, Li F, Liu Z, He C, Wu J, Shi Y & Randow F (2013) Sterical hindrance promotes selectivity of the autophagy cargo receptor NDP52 for the danger receptor galectin-8 in antibacterial autophagy. *Sci Signal* **6**, ra9.
48. He L, André S, Siebert H-C, Helmholz H, Niemeyer B & Gabius H-J (2003) Detection of ligand- and solvent-induced shape alterations of cell-growth-regulatory human lectin galectin-1 in solution by small angle neutron and x-ray scattering. *Biophys J* **85**, 511-524.
49. Göhler A, André S, Kaltner H, Sauer M, Gabius H-J & Doose S (2010) Hydrodynamic properties of human adhesion/growth-regulatory galectins studied by fluorescence correlation spectroscopy. *Biophys J* **98**, 3044-3053.
50. Kaltner H, Solís D, Kopitz J, Lensch M, Lohr M, Manning JC, Mürnseer M, Schnölzer M, Siebert H-C, André S & Gabius H-J (2008) Proto-type chicken galectins revisited: characterization of a third protein with distinctive hydrodynamic behaviour and expression pattern in organs of adult animals. *Biochem J* **409**, 591-599.

51. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797.
52. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME & Visscher PM (2013) Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* **14**, 507-515.
53. Jiménez M, Sáiz JL, André S, Gabius H-J & Solís D (2005) Monomer/dimer equilibrium of the AB-type lectin from mistletoe enables combination of toxin/agglutinin activities in one protein: analysis of native and citraconylated proteins by ultracentrifugation/gel filtration and cell biological consequences of dimer destabilization. *Glycobiology* **15**, 1386-1395.
54. Kopitz J, Fik Z, André S, Smetana K Jr & Gabius H-J (2013) Single-site mutational engineering and following monoPEGylation of the human lectin galectin-2: effects on ligand binding, functional aspects, and clearance from serum. *Mol Pharmaceut* **10**, 2054-2061.
55. André S, Kaltner H, Lensch M, Russwurm R, Siebert H-C, Fallsehr C, Tajkhorshid E, Heck AJR, von Knebel Doeberitz M, Gabius H-J & Kopitz J (2005) Determination of structural and functional overlap/divergence of five proto-type galectins by analysis of the growth-regulatory interaction with ganglioside GM1 *in silico* and *in vitro* on human neuroblastoma cells. *Int J Cancer* **114**, 46-57.
56. Kopitz J, Bergmann M & Gabius H-J (2010) How adhesion/growth-regulatory galectins-1 and -3 attain cell specificity: case study defining their target on neuroblastoma cells (SK-N-MC) and marked affinity regulation by affecting microdomain organization of the membrane. *IUBMB Life* **62**, 624-628.
57. Kopitz J, von Reitzenstein C, Burchert M, Cantz M & Gabius H-J (1998) Galectin-1 is a major receptor for ganglioside GM1, a product of the growth-controlling activity of a cell surface ganglioside sialidase, on human neuroblastoma cells in culture. *J Biol Chem* **273**, 11205-11211.
58. Kopitz J, Ballikaya S, André S & Gabius H-J (2012) Ganglioside GM1/galectin-dependent growth regulation in human neuroblastoma cells: special properties of bivalent galectin-4 and significance of linker length for ligand selection. *Neurochem Res* **37**, 1267-1276.
59. Wang J, Lu ZH, Gabius H-J, Rohowsky-Kochan C, Ledeen RW & Wu G (2009) Cross-linking of GM1 ganglioside by galectin-1 mediates regulatory T cell activity involving TRPC5 channel activation: possible role in suppressing experimental autoimmune encephalomyelitis. *J Immunol* **182**, 4036-4045.
60. Wu G, Lu ZH, Gabius H-J, Ledeen RW & Bleich D (2011) Ganglioside GM1 deficiency in effector T cells from NOD mice induces resistance to regulatory T-cell suppression. *Diabetes* **60**, 2341-2349.
61. Smetana K Jr, André S, Kaltner H, Kopitz J & Gabius H-J (2013) Context-dependent multifunctionality of galectin-1: a challenge for defining the lectin as therapeutic target. *Expert Opin Ther Targets* **17**, 379-392.

62. Thurston TLM, Wandel MP, von Muhlinen N, Foeglein A & Randow F (2012) Galectin-8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. *Nature* **482**, 414-418.
63. Ferwerda B, Ferwerda G, Plantinga TS, Willment JA, van Spriel AB, Venselaar H, Elbers CC, Johnson MD, Cambi A, Huysamen C, Jacobs L, Jansen T, Verheijen K, Masthoff L, Morré SA, Vriend G, Williams DL, Perfect JR, Joosten LA, Wijmenga C, van der Meer JW, Adema GJ, Kullberg BJ, Brown GD & Netea MG (2009) Human dectin-1 deficiency and mucocutaneous fungal infections. *N Engl J Med* **361,** 1760-1767.
64. Cheong KA, Chang YS, Roh JY, Kim BJ, Kim MN, Park YM, Park HJ, Kim ND, Lee CH & Lee AY (2011) A novel function of siglec-9 A391C polymorphism on T cell receptor signaling. *Int Arch Allergy Immunol* **154**, 111-118.
65. Hummelshøj T, Munthe-Fog L, Madsen HO, Fujita T, Matsushita M & Garred P (2005) Polymorphisms in the FCN2 gene determine serum variation and function of ficolin-2. *Hum Mol Genet* **14**, 1651-1658.
66. Garlatti V, Belloy N, Martin L, Lacroix M, Matsushita M, Endo Y, Fujita T, Fontecilla-Camps JC, Arlaud GJ, Thielens NM & Gaboriaud C (2007) Structural insights into the innate immune recognition specificities of L- and H-ficolins. *EMBO J* **26**, 623-633.
67. Garred P, Honoré C, Ma YJ, Munthe-Fog L & Hummelshøj T (2009) MBL2, FCN1, FCN2 and FCN3: the genes behind the initiation of the lectin pathway of complement. *Mol Immunol* **46**, 2737-2744.
68. Rao RM, Clarke JL, Ortlepp S, Robinson MK, Landis RC & Haskard DO (2002) The S128R polymorphism of E-selectin mediates neuraminidase-resistant tethering of myeloid cells under shear flow. *Eur J Immunol* **32**, 251-60.
69. André S, Sanchez-Ruderisch H, Nakagawa H, Buchholz M, Kopitz J, Forberich P, Kemmner W, Böck C, Deguchi K, Detjen KM, Wiedenmann B, von Knebel Doeberitz M, Gress TM, Nishimura S-I, Rosewicz S & Gabius H-J (2007) Tumor suppressor $p16^{INK4a}$: modulator of glycomic profile and galectin-1 expression to increase susceptibility to carbohydrate-dependent induction of anoikis in pancreatic carcinoma cells. *FEBS J* **274**, 3233-3256.
70. Sanchez-Ruderisch H, Fischer C, Detjen KM, Welzel M, Wimmel A, Manning JC, André S & Gabius H-J (2010) Tumor suppressor $p16^{INK4a}$: downregulation of galectin-3, an endogenous competitor of the pro-anoikis effector galectin-1, in a pancreatic carcinoma model. *FEBS J* **277**, 3552-3563.
71. Amano M, Eriksson H, Manning JC, Detjen KM, André S, Nishimura S-I, Lehtiö J & Gabius H-J (2012) Tumour suppressor $p16^{INK4a}$: anoikis-favouring decrease in N/O-glycan/cell surface

sialylation by down-regulation of enzymes in sialic acid biosynthesis in tandem in a pancreatic carcinoma model. *FEBS J* **279**, 4062-4080.

72. Valenzuela HF, Pace KE, Cabrera PV, White R, Porvari K, Kaija H, Vihko P & Baum LG (2007) O-Glycosylation regulates LNCaP prostate cancer cell susceptibility to apoptosis induced by galectin-1. *Cancer Res* **67**, 6155-6162.
73. Clark MC & Baum LG (2012) T cells modulate glycans on CD43 and CD45 during development and activation, signal regulation, and survival. *Ann N Y Acad Sci* **1253**, 58-67.
74. Cabrera PV, Amano M, Mitoma J, Chan J, Said J, Fukuda M & Baum LG (2006) Haploinsufficiency of C2GnT-I glycosyltransferase renders T lymphoma cells resistant to cell death. *Blood* **108**, 2399-2406.
75. Wang L, Mitoma J, Tsuchiya N, Narita S, Horikawa Y, Habuchi T, Imai A, Ishimura H, Ohyama C, & Fukuda M (2005) An A/G polymorphism of core 2 branching enzyme gene is associated with prostate cancer. *Biochem Biophys Res Commun* **331**, 958-963.
76. Laue TM, Shah BD, Ridgeway TM & Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding SE, Rowe AJ & Horton JC, eds), pp. 90-125. Royal Society of Chemistry, Cambridge, UK.
77. Campanero-Rhodes MA, Menéndez M, Sáiz JL, Sanz L, Calvete JJ & Solís D (2006) Zinc ions induce the unfolding and self-association of boar spermadhesin PSP-I, a protein with a single CUB domain architecture, and promote its binding to heparin. *Biochemistry* **45,** 8227-8235.
78. Kabsch W (2010) XDS. *Acta Crystallogr* **D66,** 125-132.
79. CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr* **D50,** 760–763
80. Kantardjieff KA & Rupp B (2003) Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865-1871.
81. McCoy A, Grosse-Kunstleve RW, Storoni LC & Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr* **D61**, 458-464.
82. Painter J & Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr* **D62,** 439-450.
83. Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr* **D66,** 486-501.
84. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr* **D66**, 12-21.

85. DeLano WL (2002) The PyMOL Molecular Graphics System. DeLano Scientifics, Palo Alto, CA, USA.
86. Ruiz FM, Fernández IS, López-Merino L, Lagartera L, Kaltner H, Menéndez M, André S, Solís D, Gabius H-J & Romero A (2013) Fine-tuning of prototype chicken galectins: structure of CG-2 and structure-activity correlations. *Acta Crystallogr* D**69,** 1665-1676.

Figure Legends

Fig. 1. CD spectra of wild type and F19Y mutant proteins. Far-UV (*A*) and near-UV (B) CD spectra obtained for the full-length wild type hGal-8 (continuous lines), its F19Y mutant (dash-dot lines), and for the respective wild type (dash lines) and mutant (dot lines) N-domains. Ellipticity values are normalized per mole of molecule. Spectra were recorded at 20 ºC for 0.2 mg/ml (*A*) or 1 mg/ml (*B*) protein solutions in PBS_{β}. *Inset* in *A*, changes with temperature in the ellipticity at 218 nm of the full-length wild type (squares) and F19Y mutant (circles) proteins in the absence (open symbols) and presence (filled symbols) of 0.1 M lactose. The continuous lines correspond to the fit of a sigmoidal function to experimental data (see Experimental procedures).

Fig. 2. Effect of ligand binding on the near-UV CD spectra of wild type and F19Y proteins. Spectra of 1 mg/ml solutions in PBS_{β} of full-length wild type hGal-8 (*A*), its F19Y mutant (*B*), and the respective wild type (*C*) and mutant (*D*) N-domains were recorded at 20 ºC in the absence (solid lines) and presence of 0.1 M lactose (dash lines) or 1 mM 3'-sialyllactose (dot lines).

Fig. 3. Crystallographic structure of the N-domain of wild type hGal-8. *A*, ribbon model of the domain in complex with lactose**,** showing the two sets of antiparallel β-strands (F0-F5 and S1-S6) that form the β-sandwich motif. *B*, close view of F0 and F1 β-strands highlighting interactions shorter than 3 Å which stabilize the N-terminal F0 β-strand. *C*, architecture of the carbohydrate-binding site, and *D*, superposition of our structure (light grey) with known structures of the N-domain in complex with lactose (dark green, PDB 3AP5 (43)) and in ligand-free form (light green, PDB 3AP4 (43)).

Fig. 4. Comparison of crystallographic structures of the F19Y mutant and wild type N-domains of hGal-8. *A*, ribbon model of the F19Y mutant (yellow) in complex with lactose (A chain) superposed to the structure of the respective complex of the wild type protein (blue). *B*, comparison of the carbohydrate-binding sites of the F19Y mutant (C atoms in yellow) and wild type (C atoms in light grey) N-domains. *C*, detailed comparison of the relative positioning of F0 and S1 strands in the F19Y mutant (in yellow) and wild type N-domains (in light grey)**.** Residue 19 is labeled with an asterisk, and water molecules in the structure of the mutant protein are shown as pink spheres.

Fig. 5. Calorimetric titration of wild type and F19Y mutant proteins with lactose and 3'-sialyllactose. *A*, representative plots of the heat released per mole of lactose injected during the titration of wild type (□) and F19Y (○) hGal-8 as a function of the sugar/protein molar ratio. *B*, titration of wild type hGal-8 (□) and its N-terminal domain (△) with 3'-sialyllactose. Solid lines correspond to the best fit of the experimental data using two-sets-of-sites (*A*) and one-set-of-sites (*B*) models.

Fig. 6. Binding analysis of ^{I25}I-labeled proteins to human neuroblastoma cells. Carbohydrate-dependent binding of labeled wild type (circles, solid/short dash lines) and variant (triangles, solid/long dash lines) proteins was measured in series with increasing galectin concentration using cells cultured for five days (final density: 10^5 cells per well), in the absence (filled symbols, solid lines) or presence of 250 µg/ml cholera toxin B-subunit (open symbols, dash lines). Results are the means of three independent experiments. Inset: Scatchard analysis of galectin binding. Measurements were performed at 37 ºC.

Fig. 7. Effect of lectins on neuroblastoma cell growth. Proliferation was measured after 48 h incubation in the presence of wild type (black bar) and variant (grey bar) proteins, using untreated cultures (white bar), set to 100%, in parallel as control and reference for statistical data processing (*A*). Results of calculating statistical significance of differences between the two data sets in direct comparison between wild type and mutant proteins are given in (*B*). Presented data are the means of four independent experiments ± SD, level of significance is graded into three categories: * $0.05 \geq p > 0.01$, ** $0.01 \geq p \geq 0.005$ and *** $p < 0.005$.

Fig. 8. Effect of lectins on erythroleukemia cell growth. Proliferation was measured after indicated periods of time in the presence of given concentrations of wild type (black bar) and variant (grey bar) proteins using untreated cultures in parallel as control, set to 100%, and as reference (*A*). Statistical evaluation of the two data sets in direct comparison is given in (*B*). Results are the means of five independent experiments ± SD and two independently prepared protein batches. For grading of statistical significance, please see legend to Fig. 7.

Fig. 9. Cytofluorometric analysis of galectin binding to human erythroleukemia cells and its sensitivity to lactose. Cell surface staining (percentage of positive cells/mean fluorescence intensity) by labeled wild type (*A*) and F19Y mutant (*B*) proteins at 2 µg/ml and sensitivity to presence of 1 mM, 5 mM, 10 mM and 20 mM lactose (grey area: background in the absence of lectin), all numbers given from bottom (no lactose added) to top (highest lactose concentration at 20 mM and finally the background control). Aliquots of cell suspensions from the same batch were tested in parallel in three independent series with SD-values not exceeding 12.4% after data normalization.

Fig. 10. Effect of lectins on colon adenocarcinoma cell growth. Proliferation of cells of four lines derived from microsatellite-instable (Colo60H, HCT116) or microsatellite-stable (SW480, SW707) tumors was measured after 48 h incubation in the presence of wild type (black bar) and variant (grey bar) proteins, using untreated cultures in parallel as control and as reference, set to 100% (*A*). Statistical evaluation of the two data sets in direct comparison is given in (*B*). Results are the means of four independent experiments ± SD. For grading by statistical significance, please see legend to Fig. 7.

Table 1. Parameters of wild type version and F19Y variant of full-length hGal-8 and separate N-domains. Thermal stability in the absence and presence of ligands was analyzed using circular dichroism. $^{+}$ M_w, weight-average molecular mass; * M_{app}, apparent molecular mass.

	Ligand	Full-length proteins		N-domains	
		Wild type	F19Y	Wild type	F19Y
Predicted *M* (Da)	–	35,676.9	35,692.4	18,063.1	18,079.3
Mass spectrometry					
M (kDa)	–	36.00 ± 0.05	36.0 ± 0.1	18.2 ± 0.1	18.19 ± 0.07
Sedimentation equilibrium					
M_w^{+} (kDa)	–	33± 2	36 ± 1	22.5 ± 1.5	20.5 ± 0.7
	0.1 M lactose	33 ± 1	36.0 ± 1.5	20± 2	20.5 ± 2.1
Sedimentation velocity					
$s^0_{20,w}$ (S)	–	2.7 ± 0.2	2.8 ± 0.1	2.0 ± 0.1	2.1 ± 0.2
	0.1 M lactose	2.6 ± 0.1	2.75 ± 0.35	2.1 ± 0.1	2.1 ± 0.1
f/f_0	–	1.37	1.34	1.12	1.09
	0.1 M lactose	1.42	1.36	1.07	1.04
Gel filtration chromatography					
Elution time (min)	0.1 M lactose	27 ± 0.2	27 ± 0.2	29.5 ± 0.1	29.5 ± 0.1
M_{app}^{*} (kDa)	0.1 M lactose	34.5 ± 1.5	34.5 ± 1.5	18.5 ± 0.5	18.5 ± 0.5
Thermal stability					
$T_{1/2}$ (ºC)	–	51.4 ± 0.1	53.9 ± 0.1	53.4 ± 0.1	53 ± 0.1
	0.1 M lactose	60.6 ± 0.1	61.3 ± 0.1	62.5 ± 0.1	62.5 ± 0.1
	1 mM 3'-sialyllactose	55.6 ± 0.1	58.3 ± 0.1	64.0 ± 0.1	63.7 ± 0.1

Table 2. X-Ray diffraction data and refinement statistics for the wild type version and F19Y variant of the N-terminal hGal-8 domain

	Wild type	**F19Y**
Data Collection		
Space group	$P4_32_12$	$P2_12_12_1$
Unit cell parameters (a, b, c) (Å)	49.52, 49.52, 160.50	55.36, 66.64, 83.99
Resolution range (Å)	47.31-1.35 (1.37-1.35)	46.22-1.91 (1.96-1.91)
No. of measured reflections*	344152 (16309)	94382 (6368)
No. of unique reflections*	43869 (2068)	24528 (1679)
Completeness (%)*	98.1 (93.4)	99.1 (94.3)
Rmeas (%)*	5.6 (168)	7.2 (118)
Multiplicity*	7.9 (7.9)	3.8 (3.8)
I/σ (I)*	22.1 (1.5)	8.6 (1.0)
CC (1/2)*	1.0 (0.5)	0.99 (0.6)
Mosaicity	0.15	0.21
Refinement		
R_{cryst}/R_{free}	15.8/17.7	22.9/28.3
Ramachandran outliers (%)	0.0	0.0
Ramachandran favoured (%)	99.3	97.6
No of atoms		
Protein	1206	2388
Solvent	252	99
Lactose	23	46
Additives	1	-
Ions	2	-
RMS angles (º)	1.17	1.10
RMS bonds (Å)	0.006	0.007
PDB accession ID	4BMB	4BME

*values in parentheses are for outer-resolution shell

Table 3. ITC-derived thermodynamic parameters for the binding of lactose and 3'-sialyllactose to the wild type version and F19Y variant of full-length hGal-8 and separate N-domains

	Lactose						3'-Sialyllactose		
	$K_1 \cdot 10^{-4}$ (M^{-1})	ΔH_1 (kcal/mol)	ΔS_1 (cal/(mol·K))	K_2 (M^{-1})	ΔH_2 (kcal/mol)	ΔS_2 (cal/(mol·K))	$K_1 \cdot 10^{-5}$ (M^{-1})	ΔH_1 (kcal/mol)	ΔS_1 (cal/(mol·K))
Full-length protein									
Wild type	1.1 ± 0.3	-11.2 ± 0.9	-19 ± 2	310 ± 150	-10 ± 2	-22 ± 6	2.6 ± 0.2	-12.8 ± 0.1	-18.2 ± 0.2
F19Y	1.1 ± 0.1	-11.7 ± 0.6	-21 ± 2	460 ± 100	-10.6 ± 0.9	-24 ± 3	2.5 ± 0.5	-11.5 ± 0.1	-14.0 ± 0.7
N-domain									
Wild type	1.3 ± 0.1	-11.3 ± 0.1	-19.2 ± 0.1	-	-	-	2.1 ± 0.2	-13.1 ± 0.1	-19.7 ± 0.1
F19Y	1.3 ± 0.1	-10.2 ± 0.2	-15.5 ± 0.5	-	-	-	2.1 ± 0.2	-12.8 ± 0.1	-18.7 ± 0.1

Figure 1

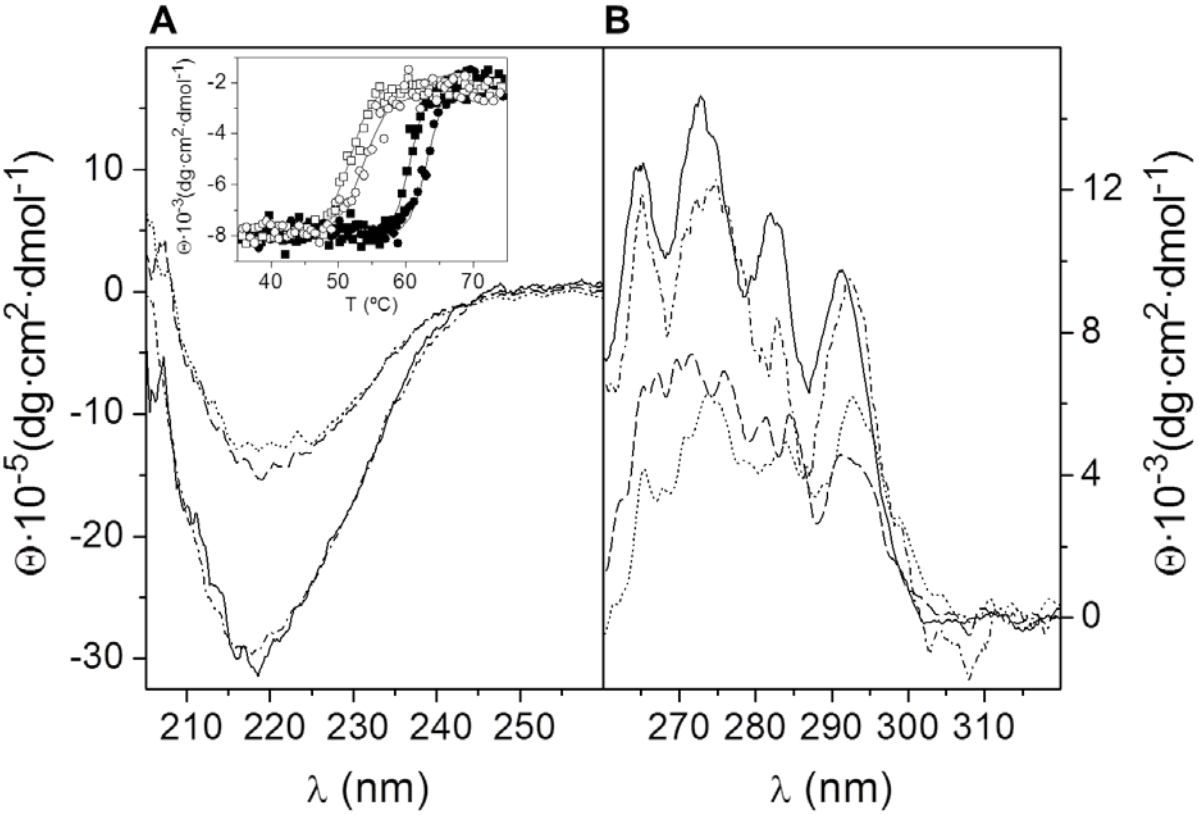

A
B
Θ·10-5(dg·cm2·dmol-1)
Θ·10-3(dg·cm2·dmol-1)
λ (nm)
λ (nm)
T (°C)
Θ·10-3(dg·cm2·dmol-1)

Figure 2

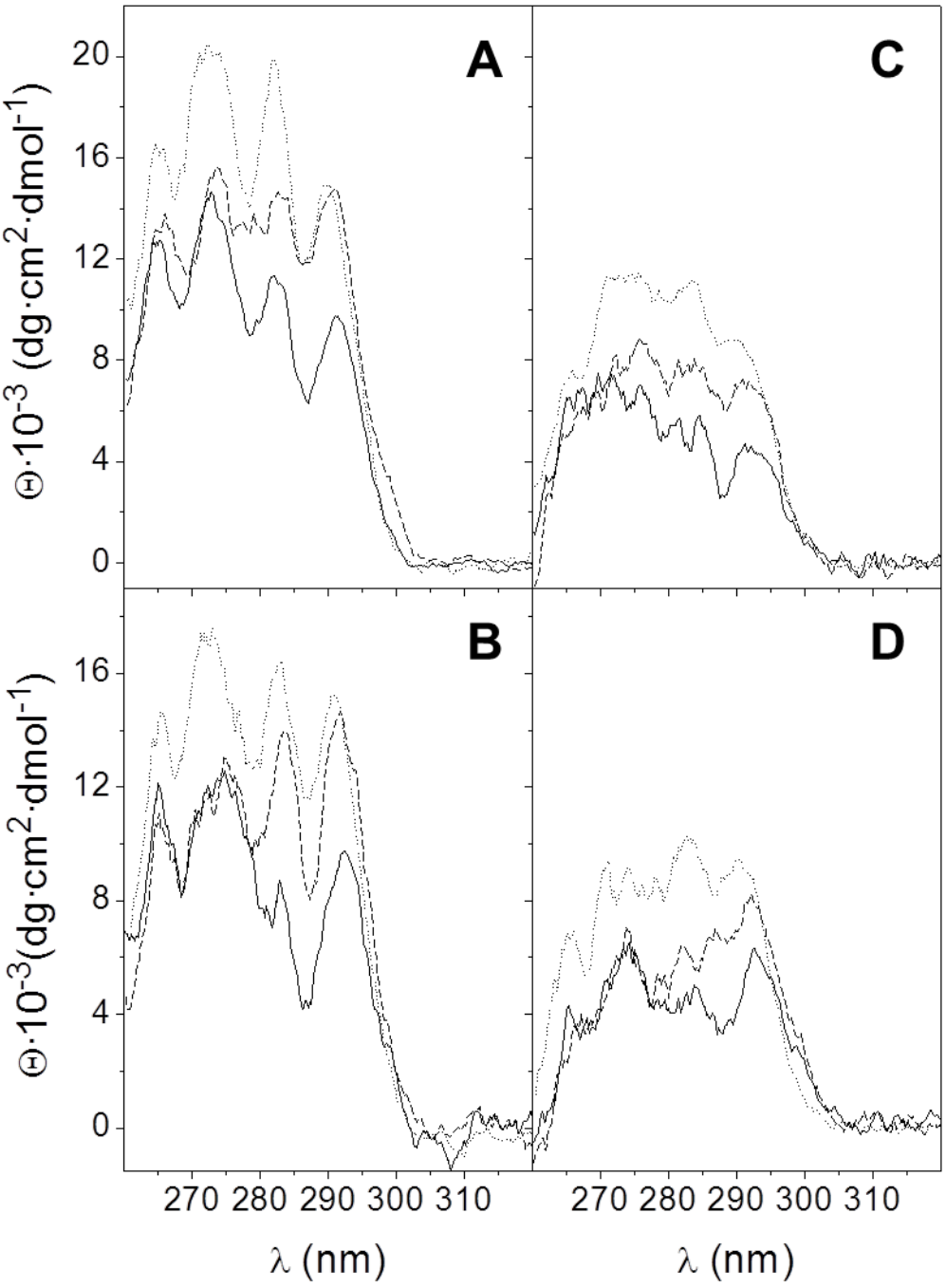

A
C
B
D
Θ·10-3 (dg·cm2·dmol-1)
Θ·10-3(dg·cm2·dmol-1)
0
4
8
12
16
20
270 280 290 300 310
λ (nm)

Figure 3

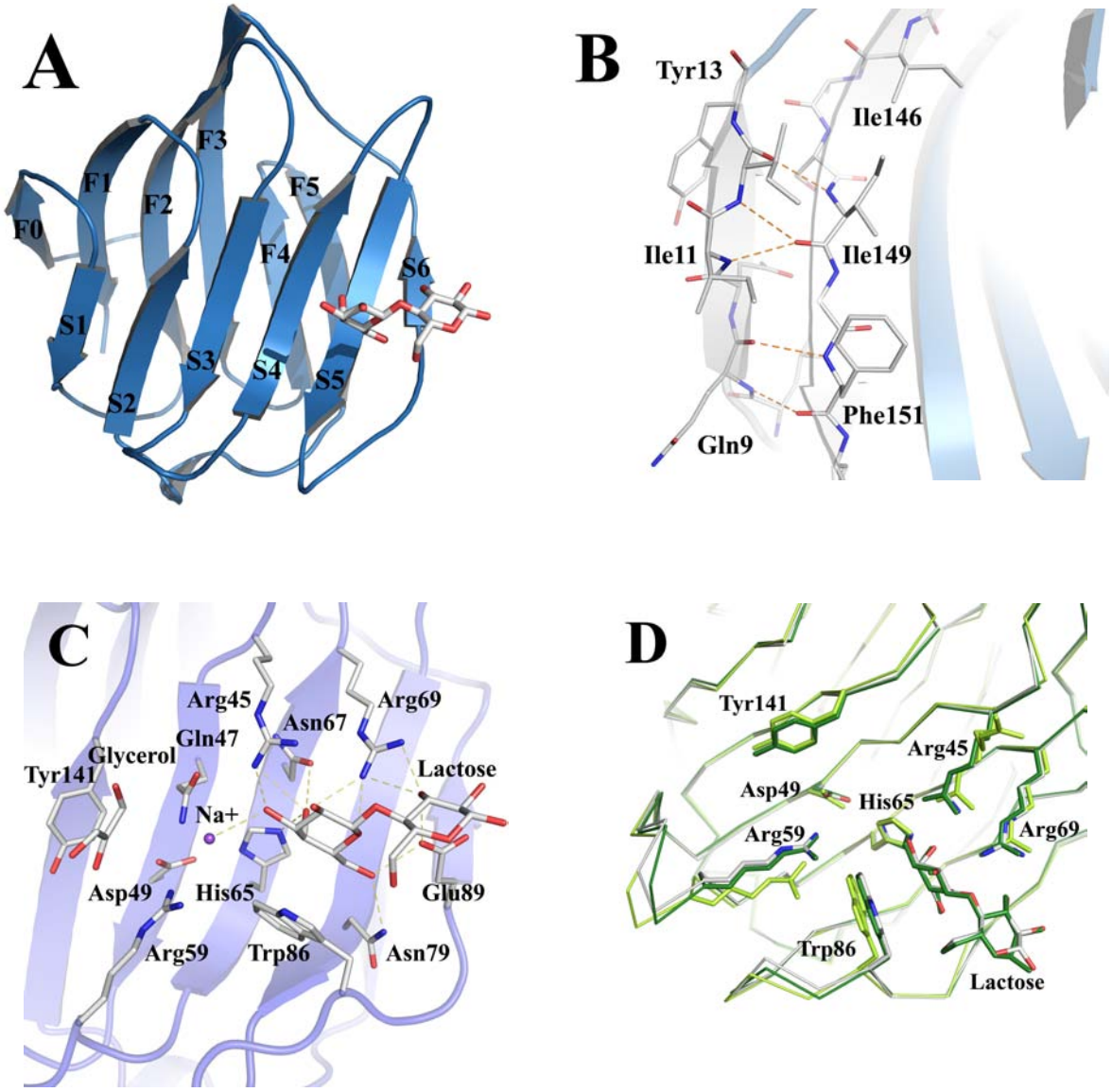

A
F0
F1
F2
F3
F4
F5
S1
S2
S3
S4
S5
S6
B
Tyr13
Ile146
Ile11
Ile149
Phe151
Gln9
C
Arg45
Gln47
Asn67
Arg69
Glycerol
Tyr141
Lactose
Na+
Asp49
His65
Glu89
Arg59
Trp86
Asn79
D
Tyr141
Arg45
Asp49
His65
Arg59
Arg69
Trp86
Lactose

Figure 4

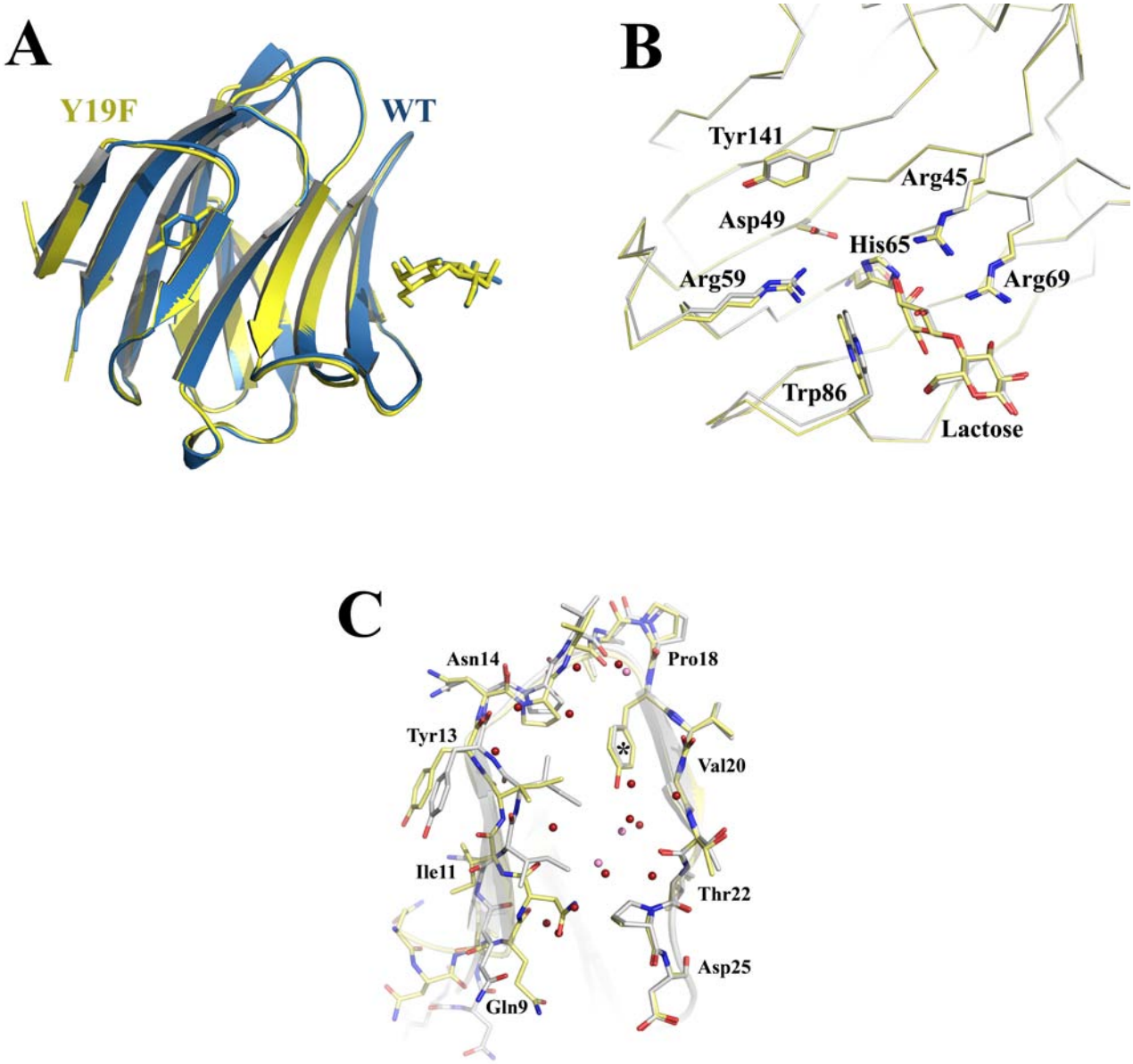

A
Y19F
WT
B
Tyr141
Arg45
Asp49
His65
Arg59
Arg69
Trp86
Lactose
C
Asn14
Pro18
Tyr13
Val20
Ile11
Thr22
Asp25
Gln9

Figure 5

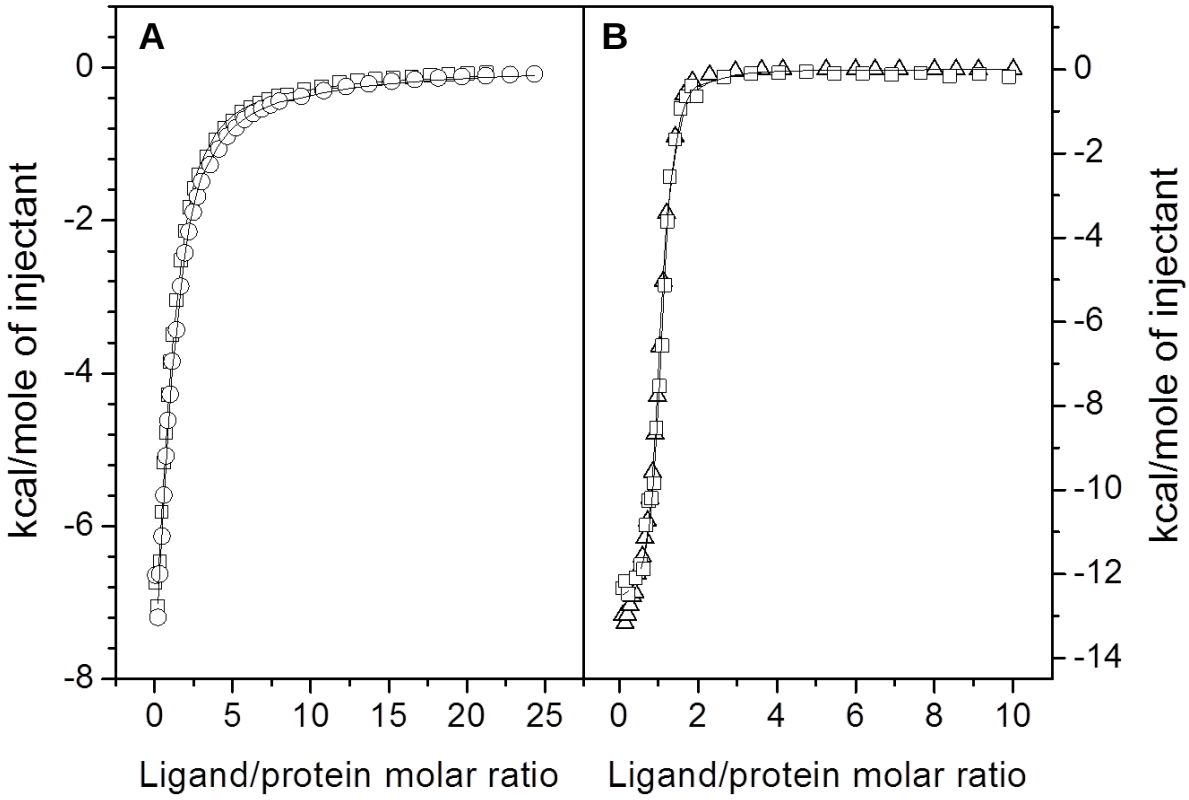

A
B
kcal/mole of injectant
kcal/mole of injectant
0
-2
-4
-6
-8
0
-2
-4
-6
-8
-10
-12
-14
0
5
10
15
20
25
0
2
4
6
8
10
Ligand/protein molar ratio
Ligand/protein molar ratio

Figure 6

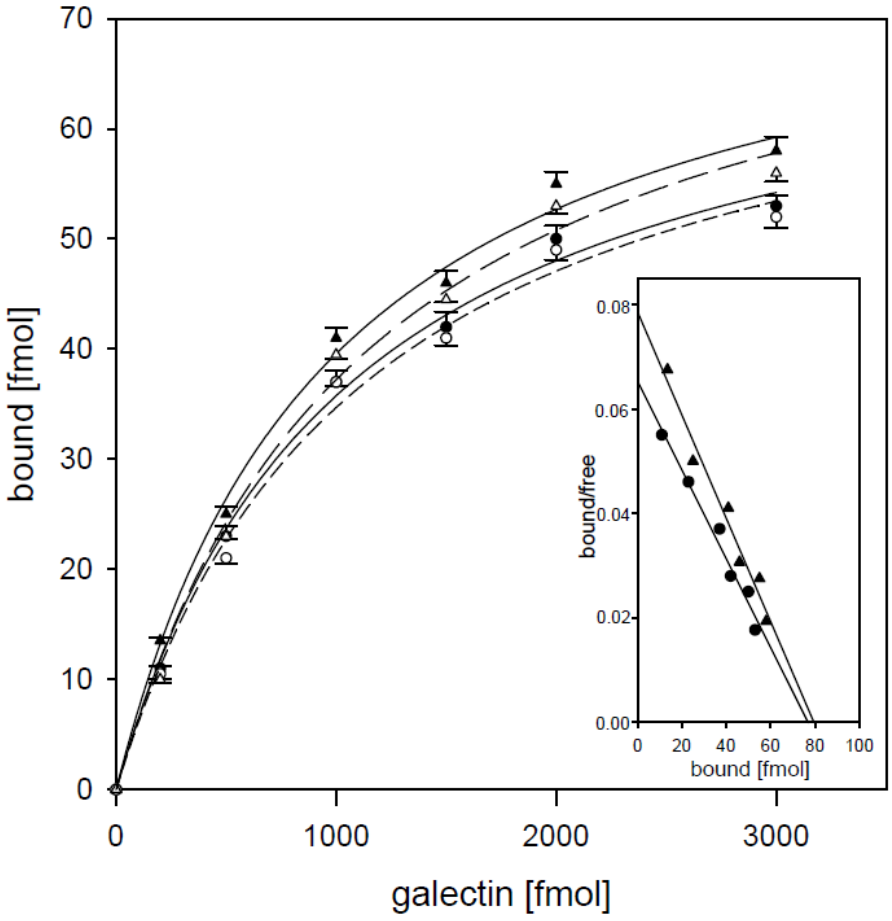


Figure 7

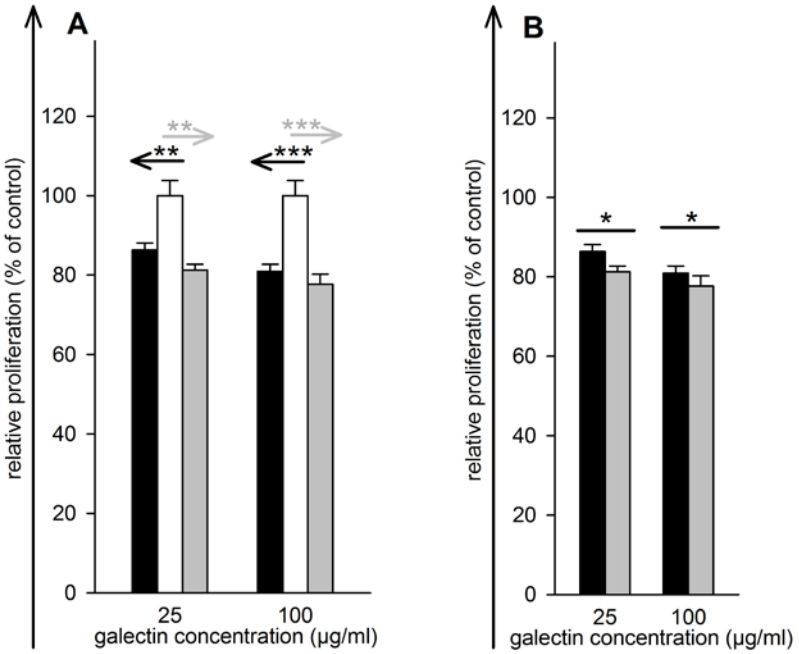

A
B
relative proliferation (% of control)
120
100
80
60
40
20
0
**
***
*
25
100
galectin concentration (µg/ml)

Figure 8

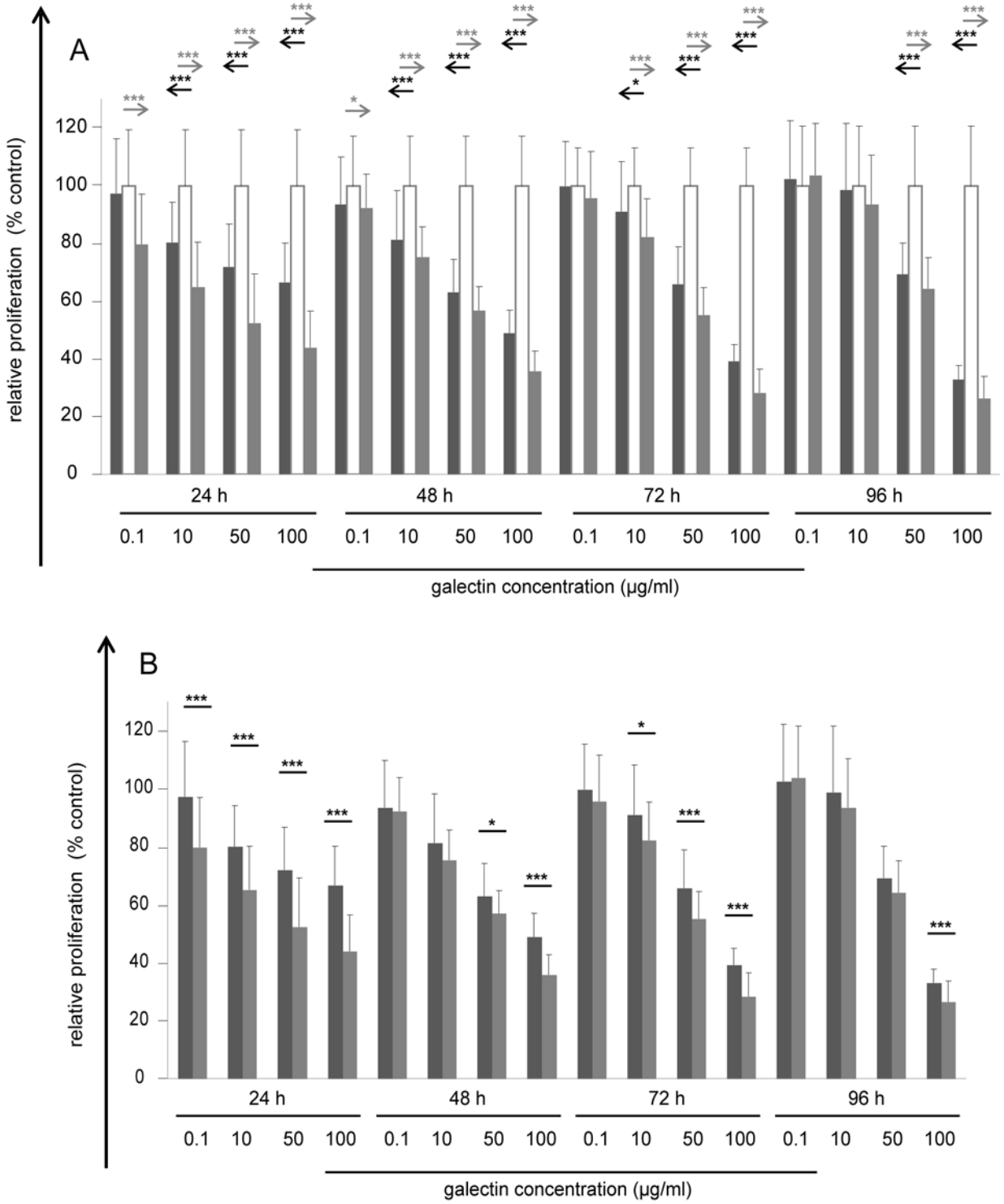

A
relative proliferation (% control)
0
20
40
60
80
100
120
24 h
48 h
72 h
96 h
0.1
10
50
100
galectin concentration (µg/ml)
B
relative proliferation (% control)
0
20
40
60
80
100
120
24 h
48 h
72 h
96 h
0.1
10
50
100
galectin concentration (µg/ml)

Figure 9

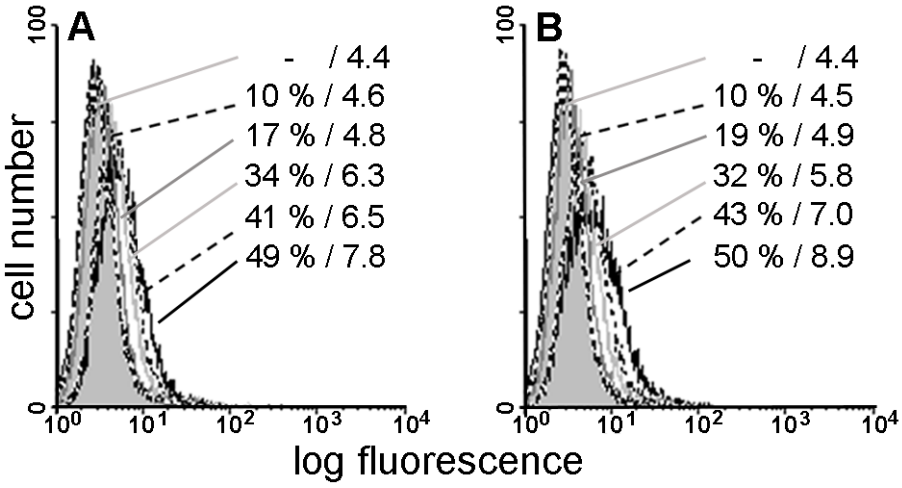
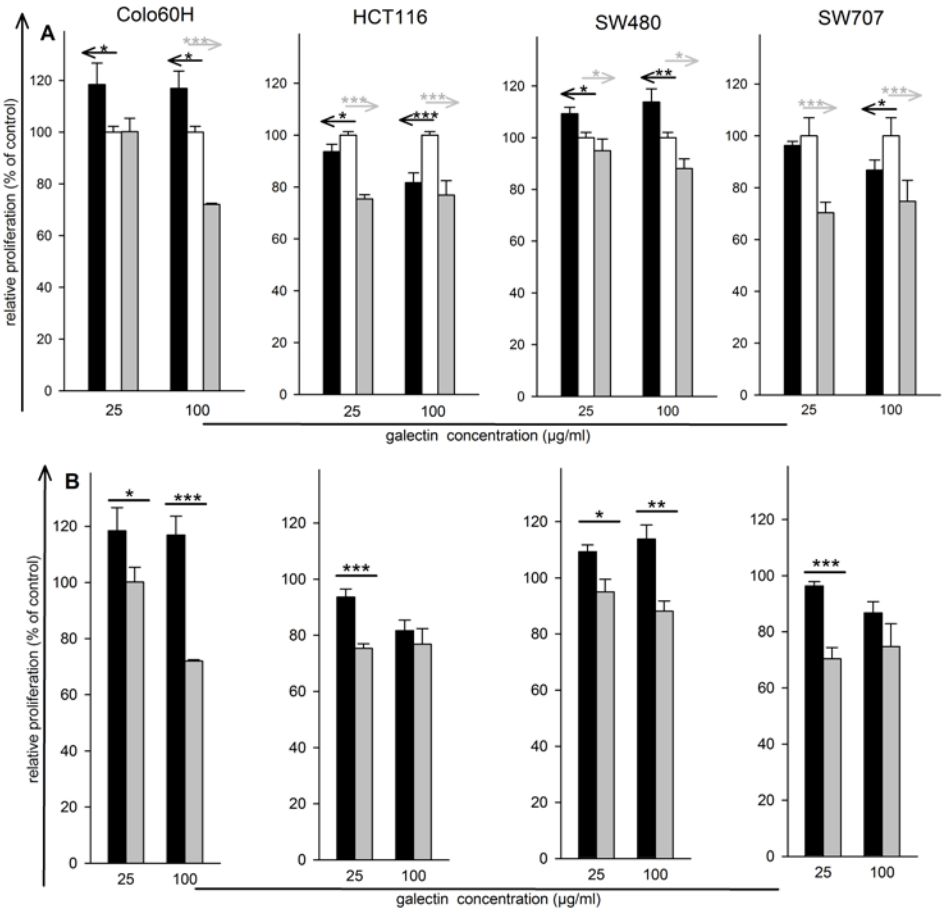

A
B
100
0
cell number
- / 4.4
10 % / 4.6
17 % / 4.8
34 % / 6.3
41 % / 6.5
49 % / 7.8
10 % / 4.5
19 % / 4.9
32 % / 5.8
43 % / 7.0
50 % / 8.9
10^0
10^1
10^2
10^3
10^4
log fluorescence

Figure 10


A
Colo60H
HCT116
SW480
SW707
relative proliferation (% of control)
galectin concentration (µg/ml)
B
relative proliferation (% of control)
galectin concentration (µg/ml)